Titre: Title:	Predicting Next Kidney Offer for a Kidney Transplant Candidate Declining Current One
Auteur: Author:	Jean-Noël Weller
Date:	2018
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Weller, J.-N. (2018). Predicting Next Kidney Offer for a Kidney Transplant Candidate Declining Current One [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/3295/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/3295/
Directeurs de recherche: Advisors:	Andrea Lodi
Programme: Program:	Maîtrise recherche en mathématiques appliquées

UNIVERSITÉ DE MONTRÉAL

PREDICTING NEXT KIDNEY OFFER FOR A KIDNEY TRANSPLANT CANDIDATE DECLINING CURRENT ONE

JEAN-NOËL WELLER
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(MATHÉMATIQUES APPLIQUÉES)
AOÛT 2018

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

PREDICTING NEXT KIDNEY OFFER FOR A KIDNEY TRANSPLANT CANDIDATE DECLINING CURRENT ONE

présenté par : WELLER Jean-Noël
en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées
a été dûment accepté par le jury d'examen constitué de :

M. ROUSSEAU Louis-Martin, Ph. D., président
M. LODI Andrea, Ph. D., membre et directeur de recherche
M. JALBERT Jonathan, Ph. D., membre

DEDICATION

« Il n'y a pas de plus grand amour
que de donner sa vie pour ses amis. »
Jean 15, 13

ACKNOWLEDGEMENTS

I wish to thank my supervisor Andrea Lodi for his confidence throughout the whole project, letting me go off the beaten methodological path. He helped me focus on realistic objectives for the project when I started losing myself into too much mathematical verbosity. He always made time when was necessary. He was always very efficient and drove the project among his impressive amount of responsibilities, with a pinch of humour.

I also thank the *Canadian National Transplant Research Program* (CNTRP) and the *Institute for Data Valorization* (IVADO) for the financial support of the research presented in this thesis.

I thank all the members of the jury for their careful proof-reading of this thesis, and their useful comments especially towards the improvement of the mathematical parts.

Je souhaite remercier le docteur Héloïse Cardinal du fond du cœur pour son engagement dans le projet, sa compréhension très large du sujet, tant sur le plan de la connaissance des reins, de l'univers de la transplantation et de l'aide à la décision que des interactions avec le milieu mathématique. Rajoutons à cela son aide attentive et dévouée, sa disponibilité, son franc-parler, sa sensibilité et son enthousiasme. Elle a été un co-superviseur officieux mais actif, et ses conseils ont été précieux dans la rédaction du présent mémoire. Sans sa ténacité, ainsi que celle de mon superviseur, l'acquisition des données aurait été proprement impossible.

J'adresse mes remerciements à Louis Beaulieu et Marie-Josée Simard de *Transplant Québec* d'avoir rendu possible l'accès aux données. Je remercie aussi Sylvain Lavigne de *Transplant Québec* et Sophie Payment de *Cosior* pour l'extraction concrète des données. Des données scientifiques telles que celles utilisées pour ce travail n'existaient pas au Québec. Merci à eux de s'être ingéniés à transformer une base de données technique en données exploitables à des fins scientifiques. Merci pour leur patience et leurs réponses à mes incompréhensions sur les données, pour le temps consacré à accéder à mes demandes sur le pointage utilisé et à comparer mes calculs avec ceux de *Transplant Québec*.

Muito obrigado à Margarida Carvalho pour ses conseils sur la rédaction scientifique. Sa disponibilité angélique l'a menée jusqu'à corriger des documents en vacances... Ses encoura-

gements ont été aussi précieux que sa connaissance de la littérature sur la transplantation rénale en mathématiques. Merci à Didier Chetelat pour ses conseils en statistique et ses remarques déstabilisantes et pointues sur mes raisonnements mathématiques. Merci à Maxime Gasse pour ses conseils sur la vérification de mes algorithmes ainsi que son approche pragmatique des mesures d'erreur et de coût. Merci à Mathieu Tanneau pour son aide, lui que j'ai fait suer du bureau jusqu'au sauna. Merci à mes voisins de bureau Pierre Hulot, Antoine Prouvost, Aurélien Serre, Flore Sentenac d'avoir à tour de rôle supporté la déblatération de mes doléances, de mes questionnements et de mes pitreries et d'avoir su quand et quand ne pas y répondre.

Merci à l'ensemble des membres de la Chaire de Recherche sur la science des données et la prise de décision en temps réel. À Koladé Nourou, pour son dévouement auprès de chacun des étudiants, son implication dans l'ensemble de la vie de la chaire, sa grande ouverture d'esprit et son combat héroïque pour maintenir les locaux salubres. À Khalid Laaziri pour son support informatique et sa bienveillance envers un amateur en console de commandes tel que moi. Il a montré qu'il était possible d'être un chercheur dans tous les domaines de la vie, de la physique à l'informatique en passant par l'histoire et la théologie. Merci à Mehdi Taobane de m'avoir aidé dans mes démarches administratives et de s'être assuré que je ne me fasse pas expulser du Canada avant la fin de ma maîtrise.

Un grand merci à l'Orchestre Symphonique des Jeunes de Montréal de m'avoir offert un espace pour souffler de tout mon cor, à tous mes collègues musiciens qui m'ont donné de la liberté vis-à-vis de mon travail scientifique, à mon professeur de cor Denys Derome, d'avoir cru qu'un étudiant en sciences pouvait continuer à se perfectionner en musique, et à mon professeur de technique Alexander Lawrence Smith, de m'avoir appris à garder la tête haute mais mobile devant mon instrument comme mon ordinateur. Merci pour leur confiance et leurs encouragements, qui, heureusement, n'ont pas suffi à me faire arrêter ma maîtrise pour me lancer complètement en musique. Merci à l'Espace Benoît Lacroix, à l'Église Unie Saint-Jean, ainsi qu'à l'Echad d'avoir donné une profondeur et une unité à mon travail pour garder à l'esprit l'essentiel de la recherche.

Merci à ma famille et mes amis, de m'avoir gardé dans leur amour, même de loin.

RÉSUMÉ

Un patient atteint d'insuffisance rénale terminale est confronté à un choix difficile lorsqu'un rein d'un donneur décédé lui est offert. Il peut ou bien accepter, ou bien attendre une meilleure offre en restant sous dialyse. La décision associe le patient et son néphrologue, qui doivent, à deux, parvenir au meilleur choix pour le patient dans un processus appelé *Prise de Décision Participative* (PDP).

D'une part, environ 500 personnes étaient en attente d'une transplantation rénale en 2017 au Québec. Parmi elles, 54 sont décédées sur la liste d'attente la même année. Le temps d'attente moyen des patients transplantés en 2017 était de 493 jours. D'autre part, des reins de moindre qualité qui auraient pu bénéficier à des patients à risque ne sont pas utilisés. Pour certains patients, accepter un rein de qualité inférieure augmente les chances de survie par rapport à attendre en dialyse. De plus, les chances de succès à long terme d'une transplantation diminuent avec le temps passé sous dialyse. Cependant, il peut être avantageux pour les patients prioritaires d'attendre une meilleur offre.

Par conséquent, une méthodologie et des outils mathématiques pour développer une PDP éclairée augmenteraient la qualité et le nombre de greffes, la survie et la satisfaction des patients, la confiance des médecins dans leurs recommandations, diminueraient les dépenses de santé et les reins non-utilisés. Toutefois, les outils qui existent à ce jour pour ce faire ne nous paraissent pas adéquats. En effet, ils visent en général à recommander au patient la décision à prendre, en se fondant sur des critères souvent incomplets, au lieu de l'informer.

Notre travail s'inscrit dans un projet de recherche plus général qui a été séparé en deux questions pour lesquelles le patient aimerait des réponses afin de prendre une décision. Notre travail répond à la seconde question en supposant la première résolue comme une boîte-noire.

1. Qu'advient-il si j'accepte l'offre ?
   Combien de temps un patient comme moi peut-il espérer survivre avec ce rein ?
   Ce temps est-il très différent du temps de survie espéré pour un donneur médian ?

2. Qu'advient-il si je refuse l'offre ?
   Combien de temps vais-je devoir attendre pour une prochaine offre ?
   Quelle est la qualité espérée d'une telle offre ?
   Combien de temps devrais-je attendre pour une meilleure offre que l'offre actuelle ?

Nous considérons un système d'attribution par pointage de reins de donneurs décédés. Des offres sont faites au patient selon son rang sur la liste d'attente, déterminé par une fonction de pointage à chaque arrivée de donneur. Soit x un patient auquel un rein y_0 est offert. Nous cherchons à prédire le temps T auquel le prochain rein Y sera proposé à x. Sachant q un prédicteur de la qualité d'une greffe (le temps de survie par exemple), nous souhaitons estimer $q(x, Y)$ ainsi que le temps de prochaine meilleure offre $(T|q(x, Y) > q(x, y_0))$.

Nous modélisons l'arrivée de donneurs éligibles (c'est-à-dire compatibles et réellement proposés au patient) par un processus ponctuel de Poisson non-homogène de paramètre constant par morceaux. En pratique, nous estimons ce paramètre à l'aide des donneurs arrivés au cours des deux années précédant l'offre. Pour chaque donneur, nous évaluons s'il aurait été éligible pour le patient en considérant différentes dates dans le futur (afin de tenir compte de l'évolution du temps d'attente et de l'âge dudit patient). En définitive, notre algorithme apprend la distribution complète de l'offre suivante pour ce patient. L'on peut fournir au patient le temps espéré $\mathbb{E}(T)$ de prochaine offre (avec intervalles de confiance obtenus par *bootstrapping*) et $t_{95\%}$, temps auquel le patient aura obtenu une offre avec probabilité 0.95.

Pour valider notre algorithme, nous utilisons des données fournies par *Transplant Québec*. Nous démontrons que comparer les quantiles prédits t_α aux temps réels de prochaine offre permet d'estimer les quantiles empiriques sur l'ensemble du jeu de données. Nous démontrons qu'il est possible de comparer la moyenne des temps observés au même mois prédit $\mathbb{E}(T)$, tout en incluant les données censurées. La meilleure version de l'algorithme prédit fidèlement la distribution de T sur l'ensemble de test (712 offres : 569 observées, 143 censurées) : temps observés inférieurs aux $t_{95\%}$ dans 94% des cas pour un C-index de 0.74. Nous introduisons une mesure de détection des erreurs de prédiction et de leur envergure. Enfin, nous utilisons le *Kidney-Donor-Risk-Index* (mesure de qualité reconnue en pratique) pour estimer la qualité de l'offre espérée comparativement à l'offre actuelle. Nous adaptons l'algorithme pour prédire l'espérance du temps de prochaine meilleure offre $\mathbb{E}(T|q(x, Y) > q(x, y_0))$.

Nous n'avons appliqué l'algorithme qu'à des données québécoises à ce jour, mais il s'étend à toute liste d'attente par pointage. Il est personnalisé, économe en temps de calcul, interprétable, s'adapte aux évolutions de la distribution de donneurs et permet d'informer le patient de multiples manières. Il inclut actuellement des limitations. Les prédictions sont mauvaises quand les données sont trop peu nombreuses ou pour certains types de patients. Enfin, l'algorithme néglige le risque de décès ou de sortie de la liste, d'où l'importance que le néphrologue confronte les résultats avec son expertise et que l'approche continue à être développée.

ABSTRACT

Patients with end-stage kidney disease waiting for a kidney transplant are confronted to difficult decisions when a deceased donor is proposed to them. They can either accept the offer, or wait for a potentially better offer, while remaining under dialysis. The decision involves both patient and physician who should evaluate together the alternative to find the best decision for the patient. This process is called *Shared-Decision-Making* (SDM).

On the one hand, around 500 persons were waiting for a kidney transplant in 2017 in the province of Québec, and 54 died in the waiting-list the same year. The mean waiting time of transplanted patients in 2017 was 493 days. On the other hand, lower-quality kidneys are wasted which could have benefited to patients at risk. For some patients, getting a lower-quality kidney leads to better survival chances than remaining on dialysis. Moreover, the longer the waiting-time, the worse the expected outcomes of a future transplant. At the same time, some high-priority patients can benefit from waiting for a better kidney offer.

Therefore, developing a methodology and decision-support tools to enhance informed SDM could at once increase quality and number of transplants, survival and satisfaction of patients, physicians' confidence in their advice, decrease organ wastage and healthcare expenditures for end-stage kidney disease. Yet, the mathematical tools which exist to foster SDM to date are not fully satisfactory. Indeed, they are designed most of the time to give an advice to the patient, based on little evidence, and not to inform him.

Our work is part of a larger research project. It has been split in two questions that the patient would like answered in order to make his decision. Our work addresses the second question and assumes that the first one is solved as a black-box.

1. What happens if I say yes?
   How long is the kidney from this specific donor expected to survive in a patient like me?
   How different is this survival compared to the one of an average donor?
2. What happens if I say no?
   If I decline this offer, how long am I supposed to wait for another offer?
   What would be the expected quality of this offer?
   How long would I have to wait for a better offer than the current one?

We consider a general scoring allocation system for deceased donors: offers are made to patients according to their rank on the waiting-list determined by a scoring function at each donor arrival. We consider a patient x getting a kidney offer y_0. Our objective is to predict the time T at which the patient will get a next offer Y. Given a black-box q predicting the quality of a matching (for example time of survival), we want to estimate the quality of next offer $q(x, Y)$ and the time to next better offer $(T|q(x, Y) > q(x, y_0))$.

The arrival of eligible donors (i.e. compatible donors who will actually be proposed to our patient) can be modelled as a non-homogeneous Poisson point process with piecewise constant parameter. We learn this parameter in practice using donors arrived up to two years before the current offer. For each donor y, we try to assess if she would have been eligible to our patient at different points in times in the future (accounting for update of patient's age and waiting-time). In the end, our algorithm predicts the whole distribution of next offer for a specific patient. This enables to provide the patient with the expected time to next offer $\mathbb{E}(T)$ (and bootstrapping derived confidence intervals) and $t_{95\%}$, time by which he will have had an offer with 95% confidence.

We validated our algorithm on data provided by *Transplant Québec*. We proved that we could compare the actual predicted quantiles t_α to the observed times to next offer and estimate the empirical quantiles over the whole dataset. We also proved that we could group the predicted expected times to next offer by month and compare them to the averaged observed times while accounting for censored values. The best version of the algorithm predicts faithfully the distribution of T on our test set (712 offers: 569 uncensored, 143 censored): actual observed times lower than predicted $t_{95\%}$ for 94% of the observations and concordance-index 0.74. We introduced a measure to detect bad predictions and study their importance. Finally we used the well-known Kidney-Donor-Risk-Index to estimate the next offer's expected quality and compare it to the current one. We adapted our algorithm to predict the mean time to next better offer $\mathbb{E}(T|q(x, Y) > q(x, y_0))$.

Though we only applied our algorithm to data from Québec, it is applicable to any scoring waiting-list. It is a highly personalised and interpretable on-line algorithm, it is not time-consuming, captures long-term trends in donors' arrival and leads to many ways of informing the patient. Of course, it currently has limitations. Bad predictions occurred for different reasons: too little data, special type of patient. Furthermore, we did not include the risk of death or removal from the list. Thus, it is important the physician should be able to confront the results to her expertise and it is also important to continue developing the approach.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ANN	Artificial Neural Network
AUROC	Area Under the Receiver Operating Characteristic curve
BMI	Body Mass Index
CDF	Cumulative Distribution Function
cPRA	calculated Panel Reactive Antibodies
CWL	Current Waiting-List
DCD	Donation after Cardiac Death
DDKT	Deceased Donor Kidney Transplantation
DND	Donation after Neurological Death
ER	Eligibility Relaxation
ESKD	End-Stage Kidney Disease
HLA	Human Leukocyte Antigen
HSP	Hypersensitised Patient
KDPI	Kidney Donor Profile Index
KDRI	Kidney Donor Risk Index
KEP	Kidney Exchange Program
KS	Kolmogorov-Smirnov
LDKT	Living Donor Kidney Transplantation
LKDPI	Living Kidney Donor Profile Index
ML	Machine-Learning
MAPE	Mean Absolute Percentage Error
MNSE	Mean Normalised Squared Error
MSE	Mean Squared Error
MWD	Mean Wasserstein Distance
NSE	Normalised Squared Error
PWL	Past Waiting-List
PP	Probability-Probability
QQ	Quantile-Quantile
SAPE	Symmetric Absolute Percentage Error
SDM	Shared Decision Making
SMAPE	Symmetric Mean Absolute Percentage Error
SRTR	Scientific Registry of Transplant Recipients
TQ	Transplant Québec

UNOS United Network for Organ Sharing

CHAPTER 1 INTRODUCTION

1.1 Renal Replacement Therapy

Kidneys are vital organs which filter blood from its impurities. The Canadian Blood Services define End-Stage Kidney Disease (ESKD) as: "A condition in which the kidneys are permanently impaired and can no longer function normally to maintain life" (CORR, 2015).

ESKD affects more than 40000 Canadians today (CORR, 2015). As population is ageing, this number is not expected to decrease in the short-term. This has an important impact on the Canadian public health system, with 1.2% of total health expenditures in 2007 (Manns et al., 2007). So far, ESKD has been addressed through renal replacement therapy, which regroups dialysis and transplantation.

The Canadian Blood Services define dialysis as a process "whereby the blood is cleaned and wastes and excess water are removed from the body" (CORR, 2015). Even though dialysis is the most easily accessible renal replacement therapy, it is less satisfactory than transplantation. In terms of quality of life, employment rate, survival expectancy and even health expenditures, the latter outperforms the former (Laupacis et al., 1996; Wolfe et al., 1999). Both therapies imply heavy medication, but dialysis usually requires three sessions of four hours per week to filter blood. Less than 50% of patients undergoing dialysis survive more than 5 years (CORR, 2015) in Canada. From an economical perspective, kidney transplantation is cost effective even for high-risk transplantations[1] in comparison to dialysis among all types of donors (Axelrod et al., 2018). Otherwise, for standard transplantations, it is cost-saving. Therefore, transplantation is indisputably the best renal replacement therapy today.

There are two types of transplantation. Living Donor Kidney Transplantation (LDKT) and Deceased Donor Kidney Transplantation (DDKT). LDKT yields better outcomes in terms of survival than DDKT (CORR, 2015). Living donations originate mostly from relatives or friends willing to give one of their two kidneys. In some cases, some unrelated altruistic donors donate a kidney. This can elicit chains of donations of willing to give but incompatible patient-donor pairs, each patient benefiting from a kidney of another donor.

[1] In the rest of this work, we call a "high-risk" donor a donor with risk features for suboptimal survival.

This happens also in so-called Kidney Exchange Programs (KEPs): willing to give but incompatible patient-donor pairs are matched to other incompatible pairs and result in cross-transplantations. However, in the majority of the cases, a patient does not have a compatible living donor, that is why around 60% of transplantations come from deceased donors in Canada (CORR, 2015).

DDKT can occur in two cases: Donation after Cardiac Death (DCD) and Donation after Neurological Death (DND). DCD happens when the donor's kidney (as well as all other relevant organs) is recovered just after the heart stops beating. DND happens when the donor's kidney is recovered while the heart keeps beating but after brain death was pronounced. In any case, the consent of the family and/or the prior consent of the donor are required. One easily understands that there is an influence of the situation in which the kidney is recovered on the expected outcomes: if the kidney is recovered after cardiac death, there is a certain time-lapse during which the organ got no oxygen. The time is very important in deceased donation: the longer an organ is outside a body (the so-called *warm ischemia time*[2]), the worse the outcomes are (Danovitch, 2009). In deceased donation, the candidates or surgeons cannot really choose the time of transplantation, unlike in living donation.

1.1.1 General Considerations about Transplant Outcomes

In general, some *sine qua non* medical conditions exist for transplantation. The transplant candidate and the donor have to be blood-type compatible: O candidates need an O donor, A candidates accept A and O, B accept B and O, and AB accept any blood-type. They also have to be *tissue-type* or *cross-match* compatible, which means the patient should not have antibodies against the donor's Human Leukocyte Antigen (HLA)[3]. To describe the probability of a patient of being tissue-type incompatible to a random donor, the cPRA is used. In brief, transplant candidates have blood tests every third month while waiting on dialysis. In a specialised laboratory, the candidate's sera are exposed to beads on which common HLA antigens are fixed. In a second step, the presence of anti-HLA antibodies is detected. The cPRA is a function of the extent to which anti-HLA antibodies are present and of the relative frequency of the HLA antigens to which a candidate has antibodies in the general population. The cPRA takes values of 0-100%. The higher the cPRA, the less likely

[2] *Warm ischemia time* is the time spent by the organ out of the body. The time spent in cold storage while the organ is transported is called *cold ischemia time*. Long cold and warm ischemia times are associated with worse graft survival.

[3] HLAs are a genetic marker of the immune footprint of a person. In transplantation, three loci are known to be particularly important: A, B and DR; the better the HLA matching between donor and patient, the better the expected survival (Danovitch, 2009).

a candidate is to find a compatible donor. It is interpreted as a patient's level of *sensitisation*. A patient is said to be *hypersensitised* if he has a greater cPRA than 80% or 95% for instance, according to the definition. This cPRA can vary over time, for example after a transfusion, a transplant or a pregnancy.

Full genetic twins put aside, a graft will always be eventually rejected (Danovitch, 2009) without appropriate medication. Therefore, all transplanted patients have to undergo immunosuppressive therapies during their whole lifetime. These therapies aim at preventing the body from rejecting the kidney. Graft rejection occurs in 15-20% of patients in the first 5 years post transplant. Rejection is a diagnosis which needs to be made on a biopsy of the transplanted kidney. It is reversible in over 75% of cases by increasing the amount of immunosuppression provided. Hence, allograft rejection does not always lead to graft failure. Graft failure is defined as a transplanted kidney that has stopped to function properly. In cases of graft failure, dialysis must be resumed, a new transplantation must be provided for patients to survive. Graft failure can happen from a short time after transplantation to several years after. The objective is always to maximise the time during which a patient can live with a functioning kidney graft.

Many parameters may influence graft outcomes: some we know before transplantation, some we know after transplantation and some nobody knows. For example, the shorter the *cold ischemia time*, the better the outcomes (more than 36 hours is often not accepted) (Danovitch, 2009; Rao et al., 2009). However, this parameter is not available when the decision to transplant is made. Adherence of the patient to the immunosuppressive therapy is also very important (Danovitch, 2009) although it is very difficult to assess beforehand (or even not desirable). Other factors that have been associated with allograft survival include donor age, age matching, donor Body Mass Index (BMI), HLA matching, patient cPRA, donor serology (HIV, hepatitis...), donor diabetes, donor hypertension, patient time under dialysis, donor ethnicity, donor gender, donor history of cigarette, cocaine... (Rao et al., 2009; Massie et al., 2016).

1.1.2 Attribution of deceased donor kidneys in Québec

In Canada, the distribution of deceased organs is organised provincially. In Québec, it is under the responsibility of *Transplant Québec* to attribute organs from deceased donors. Since the 28$^{\text{th}}$ of march 2012, there is a new attribution system in Québec. The system is split in several priority layers: renal emergency, hypersensitised patients, combined organ

transplantations patients, paediatric patients, kidney-pancreas patients, general attribution.

When a donor becomes available, her kidneys are proposed to transplant candidates according to a certain procedure to the different priority levels. If one of the kidneys is accepted at any other priority level than the general waitlist, the second kidney comes down automatically to this general waitlist. For each of the patients in the general waitlist, TQ attributes a score. All patients known to be ineligible (blood-type or tissue-type ineligible) are removed from the waitlist for the specific donor who is being investigated. The patient with the higher score gets the offer, and if he declines, the next patient on the waitlist gets an offer, until the end of the waitlist, as long as a kidney remains available.

The scoring function used by TQ is based on two criteria: utility and justice. The goal is to be simultaneously fair (minimising waiting time) and efficient (maximising survival). For justice, high waiting times and hypersensitised patients get more points. For utility, young patients, better HLA match and better donor-recipient age match get more points.

1.1.3 Decision-Making

Regardless of the allocation system and the nature of transplantation, there is always the same question whether to accept or decline an offer when a patient gets one. This is an important question as donor characteristics have a significant impact on the survival of the transplanted allograft (Rao et al., 2009; Massie et al., 2016). Also, as pointed out earlier, longer waiting-time on dialysis is also associated with increased post-transplant mortality. The decision thus depends on the patient and allograft survival perspectives with a given offer, the chances of getting a better offer in the future, the survival while waiting for another offer, the chances of becoming untransplantable in the future. It also depends on how well the patient lives under dialysis and on his own values.

The notion of *informed consent* is predominant in clinical practice: the nephrologist should provide the candidate with elements of donor history that may affect allograft survival and the candidate should participate in the decision in full knowledge of the facts. There is a subtle but important difference between informed consent and Shared Decision Making (SDM). In SDM, patients do not only receive this information and decide. Patients are expected to reflect on the options, gather more information, ask questions to their physicians, put the alternatives in the light of their preferences, values and personal characteristics to finally come to a decision jointly with their physician (Edwards and Elwyn, 2009; Gillick, 2015).

This involves greater participation and responsibility on the patient side, but studies suggest that SDM improves both outcomes and satisfaction of the patient (Greenfield et al., 1985, 1988).

Achieving SDM in practice is very difficult. Not only is there a need for reliable decision-support tools to assess the expected outcomes of the transplant, but there is also a need for education to using such tools. The latter is entrusted to the clinical world, whereas the former is up to the scientific world. Nevertheless, this is the responsibility of the scientists to develop together with the clinicians a tool of practical interest. The tool should be easy to use, it should provide reliable, understandable, unbiased and personalised information to both patient and physician. It should leave space for a discussion between patient and physician. The physician should be able to critically look at the results thanks to her expertise. So typically, a black-box administering arbitrary injunctions is not what we need.

1.2 Setting and Problem Characteristics

On the one hand, in spite of the 42% increase in deceased organ donation since 2007 in Canada, there is still a graft shortage, with around 3500 persons awaiting a transplant in 2016 (CIHI, 2017). 260 of them died in 2016 on the waitlist. In the province of Québec, around 500 persons were waiting for a kidney transplant in 2017, and 54 of them died on the waiting-list the same year (Transplant Québec, 2018). The mean waiting-time of transplanted patients in 2017 was of 493 days (Transplant Québec, 2018).[4]

On the other hand, lower-quality kidneys are wasted which could have benefited to patients at risk. For some patients, getting a kidney from a high-risk donor leads to better 2-year (or 3-year) survival chances than remaining on dialysis (Bae et al., 2016; Wey et al., 2018). Indeed, not only a patient under dialysis is at risk of an event, but the longer the waiting-time, the worse the expected outcomes of a future transplant (Meier-Kriesche et al., 2000). At the same time, some high-priority patients can benefit from waiting for a better kidney offer (Wey et al., 2018).

Therefore, developing a methodology and decision-support tools to enhance informed SDM could at once:

- increase the number of transplants

[4]This does not take into account the waiting-time of untransplanted patients during the same year.

- increase the survival of patients
- increase the satisfaction of patients
- increase the physicians' confidence
- decrease organ wastage
- decrease the cost of healthcare for ESKD

Yet, the mathematical tools which exist to foster SDM to date are not fully satisfactory. Indeed, they are designed most of the time to give an advice to the patient, and not to inform him. We think that the statistical information provided by some mathematical tools is hard to interpret for a patient, as it is given in terms of survival probability and not predicted time of survival (Wey et al., 2018). Even if we wanted to apply the methodology of, for example, Wey et al. (2018) in the province of Québec, this would not be possible due to the small size of the pool of patients and donors and the split in the data collection between post-transplant and pre-transplant. There can even be perverse side effects of too granular predictive tools with regard to the actual eventual survival (Bae et al., 2016). This problem is of course very difficult, and there will never exist a totally unbiased and at the same time understandable tool to enhance decision-making. However, we feel that work can be done in the way of extracting the information from data to foster SDM while preserving a relative neutrality.

The work presented in this thesis is part of a larger research project which aims at addressing the aforementioned problem. This large research project has been split in two questions that the patient would like to answer in order to make his decision, while complying with the specifications for the tool (easy to use, reliable, understandable, unbiased and personalised).

1. What happens if I say yes?
   How long is the kidney from this specific donor expected to survive in a patient like me?
   In someone like me, how different is the expected kidney survival provided by the current offer compared to that of an average donor, or from the best donor I could get?
2. What happens if I say no?
   If I decline this offer, how long am I supposed to wait for another offer?
   What would be the expected quality of this offer?
   How long would I have to wait for a better offer than the current one?

The second question of course relies on the first one. This thesis focuses on the second question and assumes that the first one is solved as a black-box, even though we will explore the different solutions available in the literature.

1.3 Objective

Let us consider a non-paediatric patient x in the scoring waiting-list getting a kidney offer y_0 at time 0. Our objective is to predict the time T at which the patient will get a next offer Y, provided the patient keeps being on the waiting-list in the mean-time.

Let us assume that we have a black-box q predicting the quality of an offer (for example time of survival). We want to give an estimate of the quality $q(x, Y)$ of the next offer. Then, we want to predict the time to next better offer $(T|q(x, Y) > q(x, y_0))$.

We will address this problem both mathematically, using stochastic processes, and practically, developing an algorithm which we will test on data provided by TQ.

We feel useful to mention that we will not do optimisation as such here: we do not want to find an optimal allocation model for TQ. We do not want to predict the probability that a patient dies or becomes untransplantable on the waiting-list. We do not want to train a machine learning algorithm to predict the times of interest. Neither do we want to resort to brute force simulations. Instead, we propose an original and interpretable method to give highly personalised predictions, keeping always in mind that the final objective is to enable SDM.

1.4 Outline

We will begin this work by a large literature review on the field of kidney transplantation and decision-aid, including prediction of graft survival. Then we will detail the kidney attribution system in the province of Québec: interactions between the priority lists, exceptions and the behaviour of the scoring function used to rank patients. In the next chapter, we will present our mathematical modelling of the problem: the overall stochastic process with arrivals of donors, the particular point of view of a patient getting an offer, practical algorithms to learn the parameters necessary to make predictions. We will introduce the data which we used from TQ, give the characteristics of the study population, explain how we preprocessed it, verify the first main assumptions which we did in the mathematical part. Once this is

done, we will see how to verify and compare our methods with the data, developing specific methodologies for the problem and presenting the results. In conclusion, we will explain how the generic algorithm we developed works in practice, present its limitations and mention future research directions.

CHAPTER 2 LITERATURE REVIEW

2.1 Predicting Survival after Kidney Transplantation

In this section, we will discuss the predictive tools which have been developed previously to answer the first relevant question of our decision-aid: "How long is the current offer expected to function in someone like me, and how is that different from the average donor or the best donor I could get?". We will discuss the limitations of the current approaches.

2.1.1 Models Based on Cox Regression Analysis

Although many models have been published, the only one that is widely used clinically is the Kidney Donor Profile Index (KDPI), developed by Rao et al. (2009). It relies on the Kidney Donor Risk Index (KDRI) which is a measure of the relative quality of a deceased donor kidney in comparison to a certain donor of reference, using Cox regression to model the risk of death or graft failure. The KDPI gives a measure in percentile of how much better a graft is, given the donor's parameters, in comparison to the pool of deceased donor kidneys of the year before. However, Bennett and McEvoy (2011) criticise the KDRI, as the most important parameters for graft survival: diabetes, serum creatinine, cerebrovascular cause of death and hypertension, are presented sparsely in the data. They question both the relevance and the reliability of the data: how it is measured, how rigorously it is written down. In the same spirit of the KDPI, a Living Kidney Donor Profile Index (LKDPI) has been developed by Massie et al. (2016). They extended the KDPI to living donors also, still using Cox regression, allowing for a comparison between outstanding deceased donor kidneys and poor living donor kidneys.

Although both the deceased and the living donor KDPIs are adjusted for recipient characteristics, they do not account for interactions between donor and recipient characteristics in survival predictions. The presence of such interactions has been reported by Heaphy et al. (2013), who show that the negative impact of high KDPI kidneys on 5-year graft survival is much lower when the recipients have risk features for suboptimal graft survival (i.e. older recipients, diabetics, black race). The last calculator using Cox regression to date is the one developed by Ashby et al. (2017). They built a web-based calculator to improve the selection of living donors when multiple donors are available. Unlike the two former indexes, it uses both the parameters of the patient and the donor to make predictions, including both living

and deceased donors. Moreover, it gives a probability of 5 and 10 years failure. These Cox models are a useful tool to analyse the importance of each parameter. However, Cox models are not meant to make personalised survival prediction and so they should be used carefully in decision-aid.

2.1.2 Models Based on Machine-Learning

Some patient-donor approaches have been developed in the field of Machine Learning (ML), not only for kidney transplantation, but also liver and heart transplantation. The first tackling the problem of long-term graft survival prediction are Krikov et al. (2007), with an approach based on decision trees for kidneys. They trained decision trees to predict 1, 3, 5, 7 and 10-year survival probability. They used a dataset composed of approximately 90000 transplants from the United Network for Organ Sharing (UNOS) and the Scientific Registry of Transplant Recipients (SRTR). They managed to include both censored and uncensored patients in a first step to analyse the important features, but they only kept the uncensored patients to make their predictions. Their results were measured in terms of Area Under the Receiver Operating Characteristic curve (AUROC) and scaled from 0.63 for the one year prediction to 0.90 for the ten years prediction. Besides the little use of the censored values, it must be noted that no special distinction was made between living and deceased donors. Reinaldo et al. (2010) used different basic ML algorithms both in classification (survival after a pre-specified number of years) and regression (prediction of the number of months of survival). They used Brazilian data with only 106 transplantation cases and 16 features, and used 10 fold cross-validation to obtain their validation rates. They claim to have above 90% accuracy in classification for their best algorithms (decision trees). However, this result is to be set in perspective due to the very little dataset and the fact that they do not give a confusion matrix to see how the predictions perform. Additionally, the quality of the regression part is very difficult to evaluate, due to the little interpretability of the indicators they use (e.g. Mean Squared Error (MSE)). A comparable study was published the same year by Li et al. (2010) in which they use UNOS data (based on a thousand of records) to predict early graft survival after transplantation and prediction (less than 1, between 1 and 5, 5 and 10, and more than 10-year survival), based on Bayesian networks. In this case also, the classification part performs very well in terms of accuracy, sensitivity, specificity and precision (greater than 96%). The predictive accuracy was 70% with good sensitivity and precision (greater than 85%) for the 1-year prediction but worse ones for longer survival times (sensitivity and precision less than 65%). Tang et al. (2011) performed tree based classification algorithms to respectively predict 1, 3, 5, 7 and 10-year graft survival (with increasing values of AUROC for each classifier, from 0.59 for 1 year to

0.97 for 10 years, which are coherent with Krikov et al. (2007)). They used uncensored data from both living and deceased donors, and compared the outcomes between local (single-centre) data and national data in the U.S. The discrepancy between the local and national results highlights that care should be taken when applying large-scale algorithms to a local centre. Recently, Shaikhina et al. (2017) made a more specific application of ML for outcome prediction. Again, they used decision trees and random forests to a small British dataset (less than 80 grafts) for the special case of antibody incompatible patient donors pairs. The most robust model is as expected the one they get from random forests (which results in an 85% accuracy predicting early rejection). This work underlines the particular importance of the donor-specific immunoglobulin level in rejection (in addition to more commonly known parameters as HLA mismatches).

General ML algorithms were applied for other organs than kidney, for which we will only mention two recent examples, interesting to us from a methodological point of view. The first one comes from Dag et al. (2017). They handle the 1, 5 and 9-year heart transplantation survival prediction. They tackle the problem of imbalance in the data (more successes than failures) by using resampling methods. They choose to simply discard records with missing values. It is remarkable that the best performances are achieved with Logistic Regression, with AUROC values scaling from 0.62 for one year to 0.84 for nine-year prediction. Among others, their neural network is significantly outperformed by the logistic regression. Then they measure the evolution of importance of the features through time, using a weighted average of the sensitivity of the features for their different models. The second example by Pérez-Ortiz et al. (2017) tackles the liver transplantation outcome and allocation system problems through semi-supervised learning. They rebalance their dataset using both censored patients and fictional transplants (building new pairs from the already existing donors and patients), labelled from their supervised data, with nearest neighbour algorithms and refinements. The prediction was performed through support vector classifier for 3 or 12-month survival, using the rebalanced dataset, with best trade-off between overall accuracy and predictive power for the minority class obtained in the semi-supervised approach.

2.1.3 Models Based on Artificial Neural Networks

Even though Artificial Neural Networks (ANNs) are part of ML, we consider separately their use for kidney survival prediction. To the best of our knowledge, the first attempt to use ANNs for kidney transplantation was made by Brier et al. (2003), in order to predict graft-delayed function (post operative complication). Using a small, local dataset (300 ob-

servations), ANN increased sensitivity but decreased specificity when compared to logistic regression. An interesting comparison between the Cox model and neural networks was given by Akl et al. (2008), with a dataset of 1900 patients with a living-donor graft having survived at least three months, for which they predict the 5-year survival. They show that their ANN is superior to the Cox model for the sensitivity and the positive predictive value, but equivalent for the accuracy and the specificity. The most recent example to our knowledge for an attempt to use neural networks to predict kidney transplantation survival comes from our group (Luck et al., 2017). Our colleagues used data from the SRTR (around 100000 donor-patient pairs); a Cox's partial likelihood function is used as a first loss function, designed to handle ties (several patients deceased after the same period), which they combined to a ranking loss, designed to handle right-censored data. They tried to predict the number of years of survival and the survival curve for each patient. However, the outcomes, measured with the C-statistic (0.65), are similar to the ones from the Cox model in terms of clinical interpretation. If the algorithm was able to predict faithfully the survival curves of a specific patient-donor match, this would enable to predict the time of survival with confidence intervals, which could be included in a decision-aid tool.

For more methodological information about ANNs designed for survival prediction, we refer to Faraggi and Simon (1995) for a non-linear Cox model using ANNs and Ripley et al. (2004) for a more complete overview.

2.1.4 Summary

In summary, although many models have been developed to predict patient and allograft survival in kidney transplant recipients, these tools are not used in everyday clinical practice, with the exception of the KDPI. In the U.S., the KDPI is used in the allocation system where the transplant candidates with the highest probability of survival (young, non-diabetic patients with short-time on dialysis) are given points when low-KDPI kidneys are available (i.e. the kidneys that have the best profile in terms of long-term graft survival). Transplant candidates must also pre-consent to receive offers for high-KDPI ($> 85\%$) kidneys (kidneys from donors who have risk features for shorter long-term graft survival). The KDPI does not seem to outperform donor age in survival prediction in British Columbia (Rose et al., 2018) and it has not been validated in other provinces for the time being. The KDPI is not used in Canada, and it may not be appropriate: Canadian transplant recipients experience better long-term graft survival than U.S. patients, which may be due to differences in anti-rejection drug coverage (Gill and Tonelli, 2012). Although our group has previously used

ANNs with the American SRTR data, our next step is to gather regional data with the hope of developing a model with better accuracy to answer the first relevant clinical question for our decision-aid.

2.2 Decision-Aid

Giving a measure of graft quality and building a decision-aid are two different things, even though the latter relies on the former. Many papers which claim to have built a decision-aid actually help the policy-makers rather than the patients and nephrologists, because of how the information is presented. For instance, the KDPI gives an idea of the average quality of an organ with no consideration of the candidate characteristics and interactions between donor and recipent characteristics. It does not say with which confidence a lower KDPI organ would actually result in lower outcomes. Nor does it provide any information on the consequences of refusing an organ for transplantation. However, it might be used as a reference to optimise an allocation policy, as it is a good indicator in average. Yet, some of the predictive models of survival presented earlier are used in practice for decision-making. The KDPI is commonly used in the U.S. and by isolated nephrologists worldwide, and therefore a guide was created on how to use the KDPI (Procurement and Network, 2016). This question is important as shown by Ahn and Hornberger (1996) already in 1996. The perception of the importance of the transplantation outcomes varies among patients, they argue. Therefore, they use a decision-model in order to infer, from the patients preferences, which organs they should be proposed. Note that the decision of the patient is predicted in order to optimise the policy: the goal is to improve the allocation policy. It is a different problem than considering a given allocation policy and making decision-aid without changing it. Gordon et al. (2013) foster the concept of SDM for transplantation and distinguish it from informed consent, which only necessitates an approval of the patient after the therapy has been explained. SDM also implies that the patient has a role to play in the selection of his therapy at different steps of the transplantation process.

We present below different research fields which include the patient at different degrees in the attribution process. The first field considers the problem from the allocator's side: how to properly optimise the attribution system, so that the patient has the best offers. The second aims at forecasting the effect of a policy on the size of the waiting-list. The third aims at optimising the decisions of the patients. The last one aims at fostering an informed decision of the patient.

2.2.1 Optimisation of Kidney Allocation Policies

Zenios et al. (2000) built an optimised allocation policy maximising the quality of life and minimising the mean waiting-time for all 6 main demographic groups. Markov Decision Processes have been used by different researchers for different purposes for liver transplantation (Alagoz et al., 2007, 2010; Erkin et al., 2010; Sandıkçı et al., 2013). Alagoz et al. (2007, 2010) apply it to advise the patient on what he should do and to build an optimal policy, whereas Erkin et al. (2010); Sandıkçı et al. (2013) use it to guess the patient's decision, the former using inverse optimisation, and the latter with partially observable Markov Decision Processes. This is exactly the reverse of our problem: we want to predict the behaviour of the waiting-list to inform the patient on his future transplant perspectives and not predicting the decisions of patients. However, correctly predicting patients' preferences is also a means to learn what matters to patients and how to develop a decision-aid. Bertsimas et al. (2013) developed a method to design allocation policies for renal transplantation thanks to optimisation, with fairness constraints.

Harvey (2015) thoroughly modelled the current U.S. kidney allocation process to prove the need of change of policy, so as to prevent the waitlist from increasing, thanks to a discrete event model. In another work (Harvey and Thompson, 2016), she focuses on the enlisting of wealthy patients in multiple transplantation centres to maximise their chances of getting a transplant, using a discrete event model, to estimate the influence of such behaviours at the local and national level.

2.2.2 Prediction of Waiting-List Evolution

Predicting the evolution of the waiting-list is also an issue. The problem has been addressed by Abellán et al. (2004) and a discrete event simulation model is used to quantify the evolution of the kidney waiting-list in an autonomous region of Spain. The aim is to predict if the waiting-list will increase or decrease with the current policy. Yet, it does not allow to predict a patient's waiting-time. Therefore, it is useful for the policy-maker, but not for the patient.

2.2.3 Optimisation of Patients Decisions

The problem of what type of kidney a patient should accept after a certain waiting-time has been addressed in the literature as the so-called *secretary-problem*. The basic general version of this problem is called by Chun and Sumichrast (2006) the *sequential assignment problem*, in which multiple candidates arrive at random times to get different positions in a

firm. Different rewards are attributed to candidates matched with a certain position. They prove that the problem only depends on the rank between the candidates and the rank between the positions. Another version of the problem is the *multiple-stopping problem with random horizon* as described by Krasnosielska-Kobos (2015) in which for example kidneys arrive at random times and have to be accepted or rejected by patients who can be randomly removed from the waiting-list (death, or transplantation impossibility). They prove that this problem can be reduced to a multiple-stopping problem with finite-horizon and discounted reward. The last improvement in this secretary problem, applied to kidney transplantation, have been made by Bendersky and David (2016a,b). The first paper focuses more on the secretary-problem, and how to model the horizons (uniform or Erlang distribution) and find the optimal strategy with dynamic programming, whereas the second exemplifies the case of kidney transplants: how demanding regarding the quality of the kidney a patient should be as time goes. To quantify the quality of a matching they only use ABO and HLA matching, although multiple other donor and recipient factors (age, sex, body size...) influence the probability of allograft function and survival. According to their model, the patient should be less demanding on graft quality as his waiting-time increases. This approach of decision-aid through operations research is interesting for us, because of the mathematical modelling part, the terminology used and the assumptions made. However, on the methodological part, we think it does not enhance SDM at the personalised level. It would be a good model of patients' rational behaviour to make simulations of the allocation system, but not to actually say to a patient if he should accept or decline a particular offer at a certain time.

2.2.4 Prediction of Outcomes when Declining Current Offer

We see only two very recent examples in literature of actual decision-aid tools. The first has been developed by Wey et al. (2018) who designed an algorithm to give to the patient: probability of 3-year survival if he accepts the current offer, versus probability of 3-year survival if he refuses it. Their algorithm uses data from the SRTR. It uses the KDRI to predict the quality of a certain graft and uses the similarity between donors. The very positive part of their work is the consideration of the problem in terms of final outcomes taking into account probabilities of having an event while in the waiting list. However, there is still the question of the interpretability of probabilities by a patient (also subject to their predictive power, here of 0.69 in AUROC). This study gives strong motivation for accepting relatively high-risk kidneys for some patients. The second example comes from Bandi et al. (2018). Their approach is similar to ours in the objectives: estimating the waiting-time of candidates awaiting scarce resources attributed according to a priority attribution system. They design a general framework applicable to kidney queues as they exist in the U.S. They

model the problem as a multi-class (for multiple priority groups) multi-server (for multiple qualities of kidneys) problem. They include robust optimisation tools to cope with incomplete information. On synthetic experiments, their method outperforms significantly simulation both in accuracy and computation time. They test their algorithm on the American kidney allocation system with data from the SRTR. They use the KDPI to estimate the quality of a kidney and try to predict (average and percentiles) the time to next offer and to next better offer. They include patients' preferences in terms of quality, blood-type, location and rank on the waiting-list (assumed to be determined by the waiting-time). Their method only uses publicly available information and can be used by transplant centres or physicians. The average absolute error of the estimates compared to empirical times is 15%. This is interesting because the estimates are very well calibrated and work with little information. However, this method is not individually personalised, makes clusters of similar patients, which works well in the American case, because of the high importance given to waiting-time in practice. We will try to go into further personalisation, less assumptions and more on-line adaptivity in a context of full information and collaboration with TQ, in line with the situation of the province of Québec.

CHAPTER 3 KIDNEY ATTRIBUTION SYSTEM IN QUÉBEC

This section extends and details the explanation given in 1.1.2. It is summarising and analysing the attribution procedure document provided by TQ (Transplant Québec, 2016).

3.1 Overall System

This system was designed qualitatively to maximise social welfare and come to a trade-off between utility and equity, most likely influenced by other systems worldwide (in the U.S. particularly).

3.1.1 Priority Lists

The current kidney allocation system in Québec has been set up on the 28^{th} of March 2012 (see section 1.1.2). As soon as a donor becomes available, her kidneys are proposed to the different priority lists, in the following order:

1. renal emergency patients
2. Hypersensitised Patients (HSPs)
3. combined organ (including a kidney) patients: excluding kidney-pancreas
4. paediatric patients
5. kidney pancreas patients
6. general attribution

Number of offers Several subtleties govern the attribution. Even though our work focuses on the general attribution (also called scoring waiting-list or general attribution scoring list), it is important we should understand the whole process, as what happens on higher priorities has a direct impact on this general attribution. Only one kidney per list can be accepted in the five priority lists. If a kidney is accepted in one of these priority lists, the other has to be proposed to the general attribution scoring list, unless it has been accepted in the renal emergency list, in which case the second one can be accepted in the HSP list. Once a kidney has been accepted in a priority list, all the patients from the other priority lists get into the general attribution scoring list. The patients of the priority list in which the

kidney has been accepted do not get into the general scoring list, at the exception of HSP and kidney-pancreas patients. If no kidney has been accepted at the end of the round, the two kidneys are simultaneously proposed for a double-graft in the general attribution, under some conditions.

Applicability of the priority When a donor has a single kidney, no priority is given to combined organs and kidney-pancreas patients. When a donor is above 45, paediatric patients do not have a priority. To get a priority, the donor of a kidney-pancreas patient must be: less than 50, with a BMI less than 30 and non-diabetic.

Transplant Centres Even though the attribution is managed by TQ, a certain freedom is given to the transplant centres. Thus, when a kidney is considered as too bad or of no interest with respect to their patients by a centre, the centre can decline the offer for all its patients. In this case, the remaining patients of this centre are removed from the waiting-list for this very kidney.

3.1.2 General Scoring List

Eligibility

To get an offer, a patient must be at first blood-type eligible to the donor. Eligibility as defined by TQ is more restrictive than compatibility. It is a fairness constraint to prevent shortage in specific blood-types (as is common for living unrelated donation, where very few O donors are available as they were all compatible to their related patient). Blood-type eligibility as per TQ standards is presented in table 3.1. All blood-type ineligible patients are removed from the waitlist at the beginning of the attribution process.

Table 3.1 Bloodtype eligibility as defined by TQ.

Candidate	**Donor**
O	O
A	A
B	B
AB (if $cPRA = 0$)	AB
AB (if $cPRA \geq 1$)	AB, A, B

The second condition is cross-match compatibility. Thanks to the cPRA of the patient, it is possible to know in advance if the patient will have specificities against the donor. Transplant

candidates who have an anti-HLA antibody that is specific to a given donor's HLA (donor-specific antibodies) are removed from the waitlist at the beginning of the attribution process. A patient with no pre-determined anti-HLA antibody to a given donor is thus eligible to an offer. In rare occurrences, although a transplant candidate can be deemed eligible *a priori*, the final compatibility testing in which the candidate's serum is put in contact with the donor's cells will reveal the presence of an anti-HLA antibody that had not been detected. In theses cases, transplantation will not occur and the kidney will return to the attribution list for another candidate.

The third condition for a patient to be eligible to an offer is to be *active* on the waiting-list. A patient may be namely temporary or permanently removed from the waiting-list due to a health or personal reason. Transplanted patients (from any type of donation) are removed from the waiting-list. To be active on the waiting-list, a patient also needs to be thoroughly investigated for the absence of contraindications to transplantation (severe and untreatable coronary or vascular disease, active or recent cancers, severe and untreatable infections, non-adherence or unstable mental health problems). Furthermore, transplant candidates need to have severe chronic kidney disease, defined as the need for dialysis or poor kidney function.

Scoring Parameters

In order to rank the waiting-list, some parameters need to be known for both candidate and donor (if we assume that all are active and in the general attribution). Their dates of birth, their blood-type, their HLA profile at the A, B and DR loci. Information on the date a transplant candidate has been registered on the waiting-list, the date he started dialysis when applicable, and his most recent cPRA is also required. For the donor, we need the date of death. We also need the specificities of the patients used to compute his cPRA, to be able to predict the cross-match.

Waiting date The waiting-time is computed as the time elapsed between the day of the offer and the day of the first dialysis session. However, for patients enlisted before the 28^{th} of March 2012, the waiting-time can be computed as the time elapsed between the day of the offer and either the date of the first dialysis session or the date a candidate was first registered on the waitlist, whichever is longest. This exception was made to avoid disadvantaging patients who were registered on the waitlist before the change in the attribution system was implemented in 2012. Moreover, for patients who got a transplant which survived more than three months and go back to dialysis after graft failure, the waiting date is chosen as the new

date of first dialysis. It remains unchanged if the transplant lasted less than three months.

HLA matching The HLA type of a patient is the set of alleles at different loci (A, B, BW, CW, DR, DRW, DQ...). There are two alleles by locus and the score only considers A, B and DR, so six elements in total. In the methodology used by TQ, a difference in the name of the alleles is equivalent to an incompatibility: A2 and A2 are matched, but A24 and A32 are mismatched. The risk of rejection increases when the donor and recipient HLA antigens are mismatched, i.e. when the recipients recognises a donor HLA antigen as non-self. For instance, a patient with A2 A24 will have no mismatch to a donor with A24 A24. Conversely, a patient with A24 A24 would have one mismatch to a donor with A2 A24. The quality of HLA matching is usually expressed as the number of mismatches between the donor and the recipient. This can take a value between 0 and 6, where the best matched kidneys have 0 donor-recipient mismatches, and the worst matched kidneys have 6 donor-recipient mismatches.

Attribution

Once a kidney gets into the general attribution list, all blood-type and tissue-type ineligible patients are removed. The list is ranked in decreasing order, from the highest score to the lowest. The patient with the highest score gets an offer. If he refuses or accepts but is eventually cross-match positive, the kidney is proposed to the next on waiting-list. Otherwise, the patient gets transplanted.

For the rest of this work, we will use the following qualitative terminology:

Definition 3.1.1 (Low/High-Priority Patient)**.** A patient on the general attribution scoring list with a "high" (resp. "low") likelihood of getting an offer within a "short" time is called a high-priority (resp. low-priority) patient.

We should not confuse low/high-priority patients with priority patients:

Definition 3.1.2 (Priority Patient)**.** A patient with one of the five priorities of TQ is called a priority patient. Such a patient can compete on the general attribution scoring list, but might get an offer in a priority list.

3.2 Mathematics of the Scoring Function

The scoring procedure used by TQ is detailed in Transplant Québec (2016). We give a mathematical formulation of this scoring function below, as well as an analysis of its behaviour.

3.2.1 Expression of the Scoring Function

Let $HLA = \mathcal{A}^2 \times \mathcal{B}^2 \times \mathcal{DR}^2$ be the set of all possible HLA types, $\mathcal{D}_s = [2, 100] \times HLA$ (set of donors) and $\mathcal{P}_s = [18, 100] \times HLA \times [0, 100] \times \mathbb{N}$ (set of patients). The index s refers to the "scoring": we are restricting the parameters of the patients and the donors to the ones necessary to the calculation of the score.

We define $f_s : \mathcal{P}_s \times \mathcal{D}_s \to \mathbb{R}_+$, the scoring function used by *Transplant Québec*. We can write, with $f_p : \mathcal{P}_s \to \mathbb{R}_+$ relative to the patient and $f_d : \mathcal{P}_s \times \mathcal{D}_s \to \mathbb{R}_+$ relative to the patient-donor matching:

$$\forall (x, y) \in \mathcal{P}_s \times \mathcal{D}_s,\ f_s(x, y) = f_p(x) + f_d(x, y) \tag{3.1}$$

We detail below the different parameters and then the different terms of the function (which is a sum of sub-functions corresponding to different ranking criteria). Let $(x, y) \in \mathcal{P}_s \times \mathcal{D}_s$:

- x^{HLA}: patient's HLA profile.
- $x^a \in [18, 100]$: patient's age in years at two decimal places.
- $x^{wt} \in [\![0, 10]\!]$: patient's completed waiting years since first dialysis or enlisting according to the policy.
- $x^{cPRA} \in [\![0, 100]\!]$: patient's cPRA. Note that a high cPRA implies a higher probability of being incompatible with a random donor.
- $y^a \in [2, 100]$: donor's age in years at two decimal places.
- y^{HLA}: donor's HLA profile.

We split f_s into $f_p = f_a + f_{wt} + f_{cPRA}$, relative to the patient-only score and containing 3 terms, and $f_d = f_{HLA} + f_{am}$, relative to the patient-donor points and containing 2 terms. So $f_s = f_p + f_d = f_a + f_{wt} + f_{cPRA} + f_{HLA} + f_{am}$. The terms are the following:

- $f_a : [18, 100] \to [0.5, 2.78]$ is a utility term privileging young patients:
  $f_a(x^a) = \dfrac{50}{\max(x^a, 1)}$ rounded at the second decimal.

- $f_{wt} : \mathbb{N} \to [0, 18]$ is an equity term privileging long waiting-time patients:

$$\begin{aligned} f_{wt}(x^{wt}) &= \begin{cases} 0.5x^{wt}, \ x^{wt} \leq 2 \\ 2(x^{wt} - 2), \ 3 \leq x^{wt} \leq 9 \\ 18, \ x^{wt} \geq 10 \end{cases} \\ &= \Big(0.5x^{wt} + (1.5x^{wt} - 4) \times \mathbb{1}_{t \geq 3}\Big) \times \mathbb{1}_{x^{wt} \leq 9} + 18 \times \mathbb{1}_{x^{wt} \geq 10} \end{aligned}$$

- $f_{cPRA} : [\![0, 100]\!] \to \{0, 3, 8\}$ is an equity term privileging highly sensitised patients:

$$f_{cPRA}(x^{cPRA}) = \begin{cases} 0, \ x^{cPRA} \in [0, 20[\\ 3, \ x^{cPRA} \in [20, 80[\\ 8, \ x^{cPRA} \in [80, 100] \end{cases} = 3 \times \mathbb{1}_{x^{cPRA} \geq 20} + 5 \times \mathbb{1}_{x^{cPRA} \geq 80}$$

- $f_{HLA} : HLA^2 \to \{0, 1, 4, 8\}$ is a utility term privileging good HLA matches:

$$f_{HLA}(x^{HLA}, y^{HLA}) = \begin{cases} 0, \text{ if there are 2 DR mismatches} \\ 1, \text{ if there is one DR mismatch and } x \text{ is heterozygous}^1 \\ 4, \text{ if there is one DR mismatch and } x \text{ is homozygous}^2, \\ \text{or no DR mismatch but other loci mismatches} \\ 8, \ x^{HLA} = y^{HLA} \end{cases}$$

- $f_{am} : [18, 100] \times [2, 100] \to \{0, 2, 4\}$ is a utility term privileging well age-matched patients:

$$f_{am}(x^a, y^a) = \begin{cases} 0, \ |x^a - y^a| > 20 \\ 2, \ 10 < |x^a - y^a| \leq 20 \\ 4, \ |x^a - y^a| \leq 10 \end{cases} = 2 \times \mathbb{1}_{|x^a - y^a| \leq 20} + 2 \times \mathbb{1}_{|x^a - y^a| \leq 10}$$

Remark 3.2.1. The terms of the scoring function have various weighs. Some of them aim at bringing at the top of the list some hard-to-match patients (i.e. patients for whom a donor is difficult to find because of a high cPRA) or the best HLA-matched donor-recipient

[1] Two different alleles at the same locus.
[2] Two identifical alleles at the same locus.

combinations with a possibly strong impact, some others have less impact but are more likely to discriminate between "equally valuable" patients (age, age matching). In the unlikely case of a tie, the date of first dialysis is considered. The term with the potentially highest impact is the waiting-time.

3.2.2 Behaviour of the Scoring Function

Naturally, some questions appear when analysing this scoring function. As it is partly discrete it is of course not differentiable. However, we would like to know if it has certain properties (monotony) that could be helpful in our problem.

Eligibility

An important question in our problem is: what makes a patient eligible to a certain donor? We mentioned necessary conditions in section 3.1.2. But a natural question is: all else conditions fulfilled, is there a score from which the patient is sure to get an offer? This is related to the broader question: is it possible to compute the set of donors the patient is eligible to?

Definition 3.2.1. (Eligible Donor).
In all the document, we call *Eligible Donor* to a patient, a donor which would actually be proposed to this patient if the donor was released at the time we consider for the patient.

Remark 3.2.2. This notion does not make rigorous sense *per se*, because in reality the set of "eligible donors" of a patient is not deterministic: it depends on the waiting-list of the patient and on the decisions of all the better ranked patients. Yet, this inappropriate expression will be very convenient in the rest of our work and will have different rigorous definitions according to the assumptions made.

Definition 3.2.2. (Offer).
An offer (of a donor to a patient) is an ordered pair $(x, y) \in \mathcal{P}_s \times \mathcal{D}_s$.

Let $(x_0, y_0) \in \mathcal{P}_s \times \mathcal{D}_s$ be an offer. If this is an actual offer of a donor to a patient, one feels that this has a certain conditioning impact on the patient. From which score would this patient get another offer? A natural idea would be to consider the only score of reference we possess as an order of magnitude of this eligibility score: $f_s(x_0, y_0)$. Even though this might give an idea of the typical scores necessary to get an offer for this patient in practice, this is mathematically wrong, because f_s does not preserve the ranks. Mathematically:

Lemma 3.2.1. *$\exists(x_1, x_2, y_1, y_2) \in \mathcal{P}_s^2 \times \mathcal{D}_s^2$ such that :*
$f_s(x_1, y_1) > f_s(x_2, y_1)$, $f_s(x_1, y_1) \leq f_s(x_1, y_2)$ but $f_s(x_1, y_2) < f_s(x_2, y_2)$.

Proof. Let :

- $x_1 \in \mathcal{P}_s$: age : 60, HLA : A1 A1, B1 B1, DR1 DR1, waiting years : 4, cPRA : 0 %
- $x_2 \in \mathcal{P}_s$: age : 18, HLA : A1 A1, B1 B1, DR3 DR3, waiting years : 3, cPRA : 100 %
- $y_1 \in \mathcal{D}_s$: age : 60, HLA : A2 A3, B2 B3, DR1 DR2
- $y_2 \in \mathcal{D}_s$: age : 60, HLA : A2 A3, B2 B3, DR1 DR3

$f_s(x_1, y_1) = 12.83$, $f_s(x_2, y_1) = 12.78$, $f_s(x_1, y_2) = 12.83$ and $f_s(x_2, y_2) = 16.78$.

□

We can find patients such that the ranking is not conserved for two different donors. Moreover, depending on the structure of the waiting-list, it is possible that our patient gets a kidney offer whereas he is last in the waiting-list, because all the other patients rejected it and then his current score is meaningless. Therefore we will have to explore other methods to get a better criterion of eligibility in practice, which we present later on in this work.

Monotonicity

Although the current scoring function cannot be said to be monotonous, a modified version of this function could be.

Let $y_0 \in \mathcal{D}_s$ and $x : \mathbb{R}_+ \to \mathcal{P}_s, t \mapsto x_t$, where $x_t = (x_0^a + t, x_0^{HLA}, x_0^{cPRA}, x_t^{wt})$ (update of the patient in terms of age and waiting-time). We now look at: $f_x : t > 0 \mapsto f_s(x_t, y_0)$.

Our intuition is that f_x will slowly decrease between two "waiting-anniversaries" and jump at each waiting-anniversary: the patient continuously loses points by getting older and discretely earns some by getting more waiting-time. Let $t_i > 0$ be the successive waiting-anniversaries posterior to time $t_0 = 0$, $\forall i \in \mathbb{N}^*$.

Theorem 3.2.1 (Waiting Monotonicity)**.**

(i) *$\forall i \in \mathbb{N}$, f_x is decreasing in $[t_i, t_i + 1[$.*

(ii) *$\forall i \in \{0, ..., 9 - x_0^{wt}\}$, $f_x(t_i) < f_x(t_{i+1})$.*

Proof. (i) This is immediate because only $f_a(x_0^a+t)$ varies for $t \in [t_i, t_i+1[$ and it decreases.

(ii) The minimum increase in waiting score is 0.5 (if the upper waiting score limit of 10 years is not reached), the maximum decrease in age score in one year is $\frac{50}{18} - \frac{50}{19} \simeq 0.15...$

□

Remark 3.2.3. Many things could be said about this scoring function. However, the methods which we developed afterwards do not make any assumptions about the scoring function but that it is "monotonous" as we described.

CHAPTER 4 MATHEMATICAL MODELLING OF THE PROBLEM

In this chapter, we are first modelling the random arrival of donors in the "market" before modelling the arrival of the next eligible donor (definition 3.2.1) for a patient declining the current offer. The final goal is to be able to predict the time of arrival of the next offer as well as a "picture" of the typical next donor. For a summary of the algorithm including references to the important mathematical results, see section 4.4.

4.1 General Modelling of the Attribution Process

4.1.1 Notations

Notations are important and thorny in statistical problems involving time. There are observed times, predicted times, predicted distributions and random variables following these distributions, actual underlying distributions from the observed times, time variables. Throughout this work, we will try to keep the notation as clear and simple as possible.

Sets

Let $\mathcal{P}$ and $\mathcal{D}$ be the generic sets of all possible patients and donors.

Remark 4.1.1. In practice, $\mathcal{P}$ and $\mathcal{D}$ will be the sets of all possible combinations of the relevant features for patients and donors.

We can interpret $\mathcal{P}_s$ and $\mathcal{D}_s$ (defined in 3.2.1) as the spaces of donors and patients, taking only into account the parameters important to compute the scoring function of *Transplant Québec*. We proceeded to a stratification. We can link those sets to $\mathcal{P}$ and $\mathcal{D}$ mathematically as follows:

Definition 4.1.1. Let $\sim_s$ be the equivalence relation defined over $\mathcal{P}$ defined as:
$\forall (x_1, x_2) \in \mathcal{P}^2$, $x_1 \sim_s x_2$ if x_1 and x_2 have the same parameters in the scoring function (age, waiting-time, HLA type, cPRA)

Remark 4.1.2. This is obviously an equivalence relation for it is : binary, reflexive, symmetric and transitive.

We overload the $\sim_s$ sign for the donors:

Definition 4.1.2. Let $\sim_s$ be the equivalence relation defined over $\mathcal{D}$ defined as:
$\forall (y_1, y_2) \in \mathcal{D}^2$, $y_1 \sim_s y_2$ if y_1 and y_2 have the same parameters in the scoring function (age, HLA type)

We can naturally see $\mathcal{P}_s$ (resp. $\mathcal{D}_s$) as the quotient set of $\mathcal{P}$ (resp. $\mathcal{D}$) for $\sim_s$.

Remark 4.1.3. We understand that each "patient" of $\mathcal{P}_s$ (resp. $\mathcal{D}_s$) represents a "simplified" unique version of several different patients or donors, considering patients or donors only from the point of view of TQ's scoring function. $\mathcal{P}$ (resp. $\mathcal{D}$) are also "simplified" versions of a patient or a donor to fit in our problem. Above the philosophical interest, it is highlighting the attention we will have towards datasets, as they lead to different distributions of probabilities (each element of the dataset has a different probabilistic weight).

Remark 4.1.4. At some point, we will need to know if a patient and a donor are blood-type and tissue-type compatible, regardless of their score. It is in theory possible to assess if they are compatible in $\mathcal{P}$ and $\mathcal{D}$, but in practice, to know if a patient and a donor are tissue-type compatible, it requires special analyses. We call $\mathcal{P}_s^c$ (resp. $\mathcal{D}_s^c$) the sets of scoring patients (resp. donors) with blood-type (A, B, AB or O), allowing for blood-type compatibility checking.

In rest of the work, we will simply use $\mathcal{P}$ and $\mathcal{D}$ most of the time, when no confusion is possible.

Random Variables

- Let $(T_n)_{n\in\mathbb{N}}$ be the random process in $\mathbb{R}_+$ describing the random times of arrival of donors on the market from an initial time $T_0 = 0$.
- Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence random variables in $\mathcal{D}$ describing the random donors arriving at time T_n for $n \in \mathbb{N}$.

4.1.2 Random Arrival of Donors

Assumptions

For the rest of this work, we assume:

Assumption 4.1.1. *The arrival of a new kidney (or 2 simultaneous new kidneys) is punctual in time. It corresponds to the time of death of the donor. Its acceptance or rejection by patients is also punctual in time. In practice, we will have a one day accuracy.*

Assumption 4.1.2. *The distribution of arrival of new kidneys is independent of time: the* $T_{n+1} - T_n$ *are i.i.d.* $\forall n \in \mathbb{N}$.

Assumption 4.1.3. *The distribution of the type of kidneys incoming is independent of time: the* Y_n *are i.i.d.* $\forall n \in \mathbb{N}$.

Assumption 4.1.4. *The type of incoming donor is independent of the time of arrival:* Y_n *and* T_n *are independent,* $\forall n \in \mathbb{N}$.

Remark 4.1.5. We already know that some of these assumptions are inherently wrong but will not have any practical consequences. For example, it is obvious that the frequency of arrival of donors on the market varies over time as we mentioned in the introduction 1.2. However, as long as the timespan of interest is not too large, this will not be problematic.

We make a stronger assumption which we will have to verify in practice:

Assumption 4.1.5 (Poisson Arrival of Donors)**.** *We assume kidney donors are arriving on the market following a homogeneous Poisson point process of parameter* $\lambda > 0$ *at times* $(T_n)_{n\in\mathbb{N}}$*, with* $T_0 = 0$.

Remark 4.1.6. This assumption means that the $T_{n+1} - T_n$ are i.i.d. $\forall n \in \mathbb{N}$ and follow an exponential distribution of parameter λ. For a more complete definition of Poisson point processes, see Streit (2010); Méléard (2016).

Reminders and basic Results on the Poisson Process

We give below a few reminders on the Poisson process, adapted to our problem. We remind the proofs of some results, in order to recall the typical statistical proof methodology which we will use for more advanced results in the next sections.

We remind that (Méléard, 2016):

$$\lambda = \lim_{t\to+\infty}^{a.s.} \frac{N_t}{t},$$

with $N_t = \sum_{n\in\mathbb{N}^*} \mathbb{1}_{T_n\leq t}$ the number of organs arrived before time $t \geq 0$ and $a.s.$ the almost sure convergence.

We also have this result (Méléard, 2016), deriving from the central limit theorem:

$$\frac{N_t - \lambda t}{\sqrt{\lambda t}} \underset{t \to +\infty}{\overset{\mathcal{D}}{\to}} Z \sim \mathcal{N}(0,1),$$

with $\mathcal{D}$ convergence in distribution.

This gives a practical way to estimate the parameter λ of the distribution.

We also remind that the $S_n = T_n - T_{n-1}$ are i.i.d. and follow an exponential distribution $\mathcal{E}xp(\lambda)$, $\forall n \in \mathbb{N}^*$.

4.2 Modelling Next Kidney Offer

In relation to our goal of predicting the next kidney offer of a patient, we will try to model the particular distribution of eligible donors for him from his current offer.

4.2.1 Notations

We keep the notations of section 4.1.1 and add the following.

Throughout the whole document, we will call: $x_0 \in \mathcal{P}$ and $y_0 \in \mathcal{D}$ our patient in his initial state and the initial proposed donor. Most of the time we will refer to their equivalent class in $\mathcal{P}_s$ and $\mathcal{D}_s$ as x_0 and y_0, but if it happens that we have to clarify our notation, we will write x_0^s and y_0^s instead. Let $(X_t)_{t \geq 0}$ be the random process of $\mathcal{P}$ characterising the evolution of the patient with time. We have $X_0 = x_0$.

We want to know the distribution of T, the time of arrival of the next donor, and of Y, the next eligible donor.

4.2.2 Random Arrival of Eligible Donors

Assumptions

We make the following assumption, which is reasonable in practice:

Assumption 4.2.1 (Deterministic Evolution of the Patient)**.** *In the rest of this work, when we will refer to X_t as x_t, we will implicitly consider a deterministic evolution of the patient*

over time, assuming no changes in cPRA and only an increase in the age of the patient and related waiting-time.

Beware, this assumption hides a far stronger one which we want to highlight and discuss:

Assumption 4.2.2 (No Event on the Waiting-List)**.** *We assume that the patient neither dies nor gets temporarily or permanently removed from the waiting-list.*

Remark 4.2.1. This assumption was "hidden" by the fact that we did not wish to include a patient status (active or inactive) variable in $\mathcal{P}_s$. This is very important for our methodology, because we delegate to patient and physician the task of comparing the expected waiting-time to the likelihood of an event while waiting. To include this status variable, we would have had to somehow predict the hazard likelihood of a patient considering his characteristics, which is a whole research area *per se* and can be addressed in future studies.

Homogeneous Poisson Process Thinning Eligible Arrival

In this section, kidneys arrive following a Poisson process. A deterministic subset of those kidneys is eligible to our patient. We want to know the distribution of the first incoming eligible kidney. We will show that the arrival of eligible kidneys is still a Poisson process.

Let $\mathcal{D}^* \subset \mathcal{D}$. It is a generic set which symbolises a deterministic subset of eligible donors.

Assumption 4.2.3. *We assume in a first phase that the set of eligible donors is deterministic and is* $\mathcal{D}^*$*.*

Let $N = \inf\left(n \in \mathbb{N}^* \,|Y_n \in \mathcal{D}^*\right)$ be the random variable giving the index of the first donor in $\mathcal{D}^*$. We want to characterise the distribution of T_N.

Theorem 4.2.1 (Poisson Process Thinning)**.** *T_N follows an exponential distribution* $T_N \sim \mathcal{E}xp(\lambda \times \mathbb{P}(Y_1 \in \mathcal{D}^*))$.
Thus: $\mathbb{E}(T) = \dfrac{1}{\lambda \times \mathbb{P}(Y_1 \in \mathcal{D}^*)}$ *and* $Var(T) = \dfrac{1}{(\lambda \times \mathbb{P}(Y_1 \in \mathcal{D}^*))^2}$.

Proof. See for example Streit (2010) for a proof.

We recognise the characteristic function of an exponential distribution $\mathcal{E}xp(\lambda\mu)$. □

Remark 4.2.2. It means that the arrival of eligible kidneys actually follows a Poisson distribution $\mathcal{P}(\lambda \times \mathbb{P}(Y_1 \in \mathcal{D}^*))$.

Remark 4.2.3. In our general formalism, the results can be written:

- $(T|Y_1 \in \mathcal{D}^*) \sim \mathcal{E}xp(\lambda \times \mathbb{P}(Y_1 \in \mathcal{D}^*))$
- $\mathbb{E}(T|Y_1 \in \mathcal{D}^*) = \dfrac{1}{\lambda \times \mathbb{P}(Y_1 \in \mathcal{D}^*)}$
- $Var(T|Y_1 \in \mathcal{D}^*) = \dfrac{1}{(\lambda \times \mathbb{P}(Y_1 \in \mathcal{D}^*))^2}$

In practice, it is possible to estimate $\lambda \times \mathbb{P}(Y_1 \in \mathcal{D}^*)$ as we saw in 4.1.2:

$$\lambda \times \mathbb{P}(Y_1 \in \mathcal{D}^*) = \overset{a.s.}{\lim_{t \to +\infty}} \frac{\sum_{n \in \mathbb{N}^*} \mathbb{1}_{T_n \leq t} \mathbb{1}_{Y_n \in \mathcal{D}^*}}{t} \tag{4.1}$$

Yet, as we proved in theorem 3.2.1, the score of a deterministic patient jumps at each waiting-anniversary and slowly decreases between two waiting-anniversaries. Therefore, assuming the set of eligible donors to be deterministic is not a good assumption.

Non-Homogeneous Poisson Process Eligible Arrival

In this section, kidneys still arrive following a Poisson process. A subset of those kidneys is eligible to our patient which changes at each waiting-anniversary. We want to know the distribution of the first incoming eligible kidney. We will show that the arrival of eligible kidneys is a non-homogeneous Poisson process.

Let us not assume 4.2.3 anymore from now on. Let us make the following assumption instead:

Assumption 4.2.4. *The distribution of eligible donors $\mathcal{D}_t^*$ is constant on $[t_i, t_{i+1}[$, $\forall i \in \{0, ..., m+1\}$ with $t_0 = 0$ and $t_{m+1} = +\infty$.*

Remark 4.2.4. The t_i, $\forall i \in \mathbb{N}^*$, correspond to the patient's waiting-anniversaries. Under this assumption, we consider that the decrease in waiting score due to ageing between two waiting-anniversaries is negligible.

Notations Let $N = \inf\left(n \in \mathbb{N}^* \middle| Y_n \in \mathcal{D}_{T_n}^*\right)$ and $N_i = \inf\left(n \in \mathbb{N}^* \middle| Y_n \in \mathcal{D}_{t_i}^*\right)$. To simplify the notation we call: $T = T_N$.

Under assumption 4.2.4, we understand that the arrival of eligible donors is Poisson "piece-wise" and the T_{N_i} are exponentially distributed. We prove the following result:

Theorem 4.2.2 (Non-Homogeneous Poisson Process)**.** *Eligible donors arrive following a non-homogeneous Poisson point process.*

(i) *T has a density function:*
$g(t) = \mu_i e^{-\mu_i(t-t_i)-\sum_{j=0}^{i-1}\mu_j(t_{j+1}-t_j)}$, $\forall i \in \{0,...,m\}, \forall t \in [t_i, t_{i+1}[$,
with $\mu_j = \lambda\mathbb{P}(Y_1 \in \mathcal{D}^*_{t_j})$, $\forall j \in \{0,...,m\}$

(ii) $\mathbb{E}(T) = \sum_{i=0}^{m} \dfrac{1-e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i} e^{-\sum_{j=0}^{i-1}\mu_j(t_{j+1}-t_j)}$

(iii) *Let* $\alpha \in]0,1[$.
$\mathbb{P}(T \leq t_\alpha) = \alpha \Leftrightarrow t_\alpha = t_{i(\alpha)} - \sum_{j=0}^{i(\alpha)-1} \dfrac{\mu_j}{\mu_{i(\alpha)}}(t_{j+1}-t_j) - \dfrac{\ln(1-\alpha)}{\mu_{i(\alpha)}}$,
where $i(\alpha) = \max(i \in \{0,...,m\} | t_i \leq t_\alpha)$

Remark 4.2.5. We can better understand the expression of the expected value of T if we write it like this:

$$\mathbb{E}(T) = \sum_{i=0}^{m} \mathbb{E}(T_{N_i})\mathbb{P}(t_i \leq T \leq t_{i+1})$$

If for some i, $\mu_i = 0$, the expression is still asymptotically true because:

$$1 - e^{-\mu_i(t_{i+1}-t_i)} = \mu_i(t_{i+1}-t_i) + o(\mu_i)$$

This means we just add the elapsed time $t_{i+1} - t_i$ to the expected value.

Proof. From assumption 4.2.4, we deduce that the arrival of eligible donors is a non-homoegeneous Poisson process, by definition, with a parameter: $\mu(t) = \lambda\mathbb{P}(Y_1 \in \mathcal{D}^*_{t_i})$, $\forall t \in [t_i, t_{i+1}[$ and $\forall i \in \{0,...,m\}$. For more information on non-homogeneous Poisson processes, see Streit (2010) for example.

(i) Let us prove first the density function.

Let $i \in \{0,...,m\}$. We prove that, $\forall t \in [t_i, t_{i+1}]$:
$\mathbb{P}(T \geq t) = \prod_{j=0}^{i-1} e^{-\mu_j(t_{j+1}-t_j)} \times e^{-\mu_i(t-t_i)}$ by induction on i.

Initialisation: $\mathbb{P}(T \geq t_0) = 1$

Heredity: $i \to i+1$

$$\mathbb{P}(T \geq t | T \geq t_i) = \frac{\mathbb{P}(T \geq t)}{\mathbb{P}(T \geq t_i)}$$
$$\Leftrightarrow \mathbb{P}(T \geq t) = \mathbb{P}(T \geq t | T \geq t_i)\mathbb{P}(T \geq t_i)$$

Between time t_i and $t \leq t_{i+1}$, $\mathcal{D}_t^* = \mathcal{D}_{t_i}^*$. So $\mathbb{P}(T \geq t | T \geq t_i) = \mathbb{P}(T_i \geq t - t_i) = e^{-\mu_i(t-t_i)}$ (we are focusing on a timespan $[t_i, t]$ over which the arrival of eligible kidneys is Poisson of parameter μ_i according to theorem 4.2.1). We can conclude by simple application of the induction hypothesis.

We conclude by differentiating: $\dfrac{\partial \mathbb{P}(T \geq t)}{\partial t} = -g(t)$.

(ii) We calculate explicitly the expected value:

$$\begin{aligned}
\mathbb{E}(T) &= \int_0^{+\infty} t g(t) dt \\
&= \sum_{i=0}^{m} \int_{t_i}^{t_{i+1}} t g(t) dt \\
&= \sum_{i=0}^{m} \int_{t_i}^{t_{i+1}} t \mu_i e^{-\mu_i(t-t_i) - \sum_{j=0}^{i-1} \mu_j (t_{j+1} - t_j)} dt \\
&= \sum_{i=0}^{m} \mu_i e^{-\sum_{j=0}^{i-1} \mu_j (t_{j+1} - t_j)} \int_{t_i}^{t_{i+1}} t e^{-\mu_i(t-t_i)} dt \\
&= \sum_{i=0}^{m} e^{-\sum_{j=0}^{i-1} \mu_j (t_{j+1} - t_j)} \left(t_i - t_{i+1} e^{-\mu_i(t_{i+1} - t_i)} + \frac{1 - e^{-\mu_i(t_{i+1} - t_i)}}{\mu_i} \right) \\
&= \sum_{i=0}^{m} e^{-\sum_{j=0}^{i-1} \mu_j (t_{j+1} - t_j)} \frac{1 - e^{-\mu_i(t_{i+1} - t_i)}}{\mu_i}, \text{ by using the fact that the first half}
\end{aligned}$$

under the sum symbol is telescopic

(iii) We calculate t_α. We have to solve:

$$\begin{aligned}
&\mathbb{P}(T \leq t_\alpha) = \alpha \\
\Leftrightarrow &1 - e^{-\mu_{i(\alpha)}(t - t_{i(\alpha)}) - \sum_{j=0}^{i(\alpha)-1} \mu_j (t_{j+1} - t_j)} = \alpha, \text{ with } i(\alpha) = \max(i \in \{0, ..., m\} | t_i \leq t_\alpha) \\
\Leftrightarrow &- \mu_{i(\alpha)}(t - t_{i(\alpha)}) - \sum_{j=0}^{i(\alpha)-1} \mu_j (t_{j+1} - t_j) = \ln(1 - \alpha)
\end{aligned}$$

$$\Leftrightarrow t_\alpha = t_{i(\alpha)} - \sum_{j=0}^{i(\alpha)-1} \frac{\mu_j}{\mu_{i(\alpha)}}(t_{j+1} - t_j) - \frac{\ln(1-\alpha)}{\mu_{i(\alpha)}}$$

□

We obtain a closed formula for the variance of T, which will be useful for our verification methodology in section 6.1:

Lemma 4.2.1 (Non-Homogeneous Poisson Process Variance)**.**

$$\mathrm{Var}(T) = 2\sum_{i=0}^{m}\left(\frac{t_i - t_{i+1}e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i} + \frac{1-e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i^2}\right)e^{-\sum_{j=0}^{i-1}\mu_j(t_{j+1}-t_j)} - \mathbb{E}(T)^2$$

Remark 4.2.6. If $\mu_i = 0$, for some $i \in \{0, ..., m-1\}$, we replace $\frac{t_i - t_{i+1}e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i} + \frac{1-e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i^2}$ by $\frac{t_{i+1}^2 - t_i^2}{2}$ (which we obtain asymptotically by doing $\mu_i \to 0$ in the general expression). Indeed:

$$\begin{aligned}
\frac{t_i - t_{i+1}e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i} + \frac{1-e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i^2} &= \frac{(t_i - t_{i+1})(1 - t_{i+1}\mu_i) + o(\mu_i)}{\mu_i} \\
&\quad + \frac{\mu_i(t_{i+1}-t_i) - \frac{1}{2}\mu_i^2(t_{i+1}-t_i)^2 + o(\mu_i^2)}{\mu_i^2} \\
&= \frac{1}{\mu_i}\Big((t_i - t_{i+1})(1 - t_{i+1}\mu_i) + (t_{i+1} - t_i) \\
&\quad -\frac{1}{2}\mu_i(t_{i+1}-t_i)^2 + o(\mu_i)\Big) \\
&= -t_{i+1}(t_i - t_{i+1}) - \frac{1}{2}(t_{i+1}-t_i)^2 + o(1) \\
&= \frac{1}{2}\left(t_{i+1}^2 - t_i^2\right)
\end{aligned}$$

Proof. By definition: $\mathrm{Var}(T) = \mathbb{E}(T^2) - \mathbb{E}(T)^2$.

We only have to calculate the first term. We find:

$$\mathbb{E}(T^2) = \sum_{i=0}^{m} \mu_i e^{-\sum_{j=0}^{i-1}\mu_j(t_{j+1}-t_j)} \int_{t_i}^{t_{i+1}} t^2 e^{-\mu_i(t-t_i)}dt$$

We find after two integrations by parts:

$$\int_{t_i}^{t_{i+1}} t^2 e^{-\mu_i(t-t_i)}dt = \frac{t_i^2 - t_{i+1}^2 e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i} + \frac{2}{\mu_i}\left(\frac{t_i - t_{i+1}e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i} + \frac{1-e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i^2}\right)$$

In the end, after removing the telescopic sum which appeared:

$$\mathbb{E}(T^2) = 0 + 2\sum_{i=0}^{m}\left(\frac{t_i - t_{i+1}e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i} + \frac{1-e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i^2}\right)e^{-\sum_{j=0}^{i-1}\mu_j(t_{j+1}-t_j)}$$

□

Now, the next step is to find an estimate of the 95% confidence intervals for $\mathbb{E}(T)$, and to find how many μ_i are necessary for a good overall $\mathbb{E}(T)$ estimate. We will see it is possible to get a closed formula for the latter, but we will use bootstrapping in practice to estimate the former.

In theory, the number of μ_i to estimate can be as high as 10 (because the scoring function gives supplementary point at each supplementary waiting-year up to the 10th year). However, we can intuit that not all values are necessary to compute a faithful estimation of $\mathbb{E}(T)$. We get the following result, which can be used as an early-stopping criterion in practice when computing the μ_i:

Theorem 4.2.3 (Approximation of the Non-Homogeneous Poisson Distribution)**.**
For $m_0 \in \{0, ..., m\}$, let τ_{m_0} be the estimator of the expected time:

$$\tau_{m_0} = \sum_{i=0}^{m_0-1}\frac{1-e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i}e^{-\sum_{j=0}^{i-1}\mu_j(t_{j+1}-t_j)} + \frac{e^{-\sum_{j=0}^{m_0-1}\mu_j(t_{j+1}-t_j)}}{\mu_{m_0}}$$

Assume $\mu_{m_0} \leq \mu_i,\ \forall i \in \{m_0, ..., m\}$. Then:

(i) $\mathbb{E}(T) - \tau_{m_0} \leq 0$

(ii) $|\mathbb{E}(T) - \tau_{m_0}| \leq \dfrac{e^{-\sum_{j=0}^{m_0}\mu_j(t_{j+1}-t_j)}}{\mu_{m_0}}$

Remark 4.2.7. The expression of τ_{m_0} seems to differ from the one of $\mathbb{E}(T)$ because the latter uses the convention: $t_{m+1} = +\infty$. Secondly, the assumption $\mu_i \leq \mu_{i+1},\ \forall i \in \{0, ..., m\}$ is very reasonable as the patient-only score increases at each time of change t_i.

Proof.

$$\mathbb{E}(T) - \tau_{m_0} = \sum_{i=m_0}^{m}\frac{1-e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i}e^{-\sum_{j=0}^{i-1}\mu_j(t_{j+1}-t_j)} - \frac{e^{-\sum_{j=0}^{m_0-1}\mu_j(t_{j+1}-t_j)}}{\mu_{m_0}}$$

$$= e^{-\sum_{j=0}^{m_0-1}\mu_j(t_{j+1}-t_j)}\left(\sum_{i=m_0}^{m}\frac{1-e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i}e^{-\sum_{j=m_0}^{i-1}\mu_j(t_{j+1}-t_j)} - \frac{1}{\mu_{m_0}}\right)$$
$$= \epsilon_{m_0}A_{m_0}$$

$\epsilon_{m_0} = e^{-\sum_{j=0}^{m_0-1}\mu_j(t_{j+1}-t_j)}$ is completely known and positive (it does not use μ_i with $i > m_0$). Let us analyse:

$$\begin{aligned}
A_{m_0} &= \sum_{i=m_0}^{m}\frac{1-e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i}e^{-\sum_{j=m_0}^{i-1}\mu_j(t_{j+1}-t_j)} - \frac{1}{\mu_{m_0}} \\
&= \sum_{i=m_0+1}^{m}\frac{1-e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i}e^{-\sum_{j=m_0}^{i-1}\mu_j(t_{j+1}-t_j)} - \frac{e^{-\mu_{m_0}(t_{m_0+1}-t_{m_0})}}{\mu_{m_0}} \\
&\leq \frac{1}{\mu_{m_0}}\left(\sum_{i=m_0+1}^{m} e^{-\sum_{j=m_0}^{i-1}\mu_j(t_{j+1}-t_j)} - e^{-\sum_{j=m_0}^{i}\mu_j(t_{j+1}-t_j)} - e^{-\mu_{m_0}(t_{m_0+1}-t_{m_0})}\right), \\
&\text{as } \mu_{m_0} \leq \mu_i,\ \forall i \geq m_0 \\
&= \frac{1}{\mu_{m_0}}\left(e^{-\mu_{m_0}(t_{m_0+1}-t_{m_0})} - e^{-\sum_{j=m_0}^{m}\mu_j(t_{j+1}-t_j)} - e^{-\mu_{m_0}(t_{m_0+1}-t_{m_0})}\right), \\
&\text{as the sum is telescopic} \\
&= \frac{-e^{-\sum_{j=m_0}^{m}\mu_j(t_{j+1}-t_j)}}{\mu_{m_0}} \\
&\leq 0
\end{aligned}$$

We keep the broad inequality, as it might occur that $m_0 = m$ and in this case, we have $e^{-\infty} = 0$. On the other hand:

$$\begin{aligned}
A_{m_0} &= \sum_{i=m_0+1}^{m}\frac{1-e^{-\mu_i(t_{i+1}-t_i)}}{\mu_i}e^{-\sum_{j=m_0}^{i-1}\mu_j(t_{j+1}-t_j)} - \frac{e^{-\mu_{m_0}(t_{m_0+1}-t_{m_0})}}{\mu_{m_0}} \\
&\geq \frac{-e^{-\mu_{m_0}(t_{m_0+1}-t_{m_0})}}{\mu_{m_0}}
\end{aligned}$$

Thus, as $\epsilon_{m_0} > 0$, we get:

$$0 \geq \epsilon_{m_0}A_{m_0} = \mathbb{E}(T) - \tau_{m_0} \geq \epsilon_{m_0}\frac{-e^{-\mu_{m_0}(t_{m_0+1}-t_{m_0})}}{\mu_{m_0}} = \frac{e^{-\sum_{j=0}^{m_0}\mu_j(t_{j+1}-t_j)}}{\mu_{m_0}}$$

□

In practice, we can estimate each μ_i with equation 4.1.

4.2.3 Eligibility Relaxation

In this section, a donor is not eligible or ineligible, she has a certain probability of being eligible according to the rank of the patient in the waiting-list. We give mathematical support to this "eligibility relaxation".

Relaxation

We mentioned in equation 4.1 a way to estimate the Poisson parameter μ_i of a thinned Poisson process, by counting the number of eligible donors who are in our set over the considered timespan. However, we know that in practice, we have a finite horizon and so a finite precision. If the process is thinned a lot (very few eligible donors), we might have a very imprecise estimation of μ_i.

It is likely that a "typically low-scoring patient" would be predicted only very few eligible donors. Thus we would estimate a very little μ. Remember that the μ are estimated by counting the number of eligible donors and dividing by the time window. Then, the differences in the μ between 1 eligible donor and 2 eligible donors would cause huge differences in the predicted times to next offer. Even worse, no eligible donor would mean an infinite waiting-time (or at least a likelihood 0 of getting an offer during the amount of time the Poisson parameters are zero). This highlights the high variability in the predictions for low-scoring patients.

To tackle this issue, we will relax the counting of eligible donors. Instead of having a discrete eligibility $\{0, 1\}$, we have a continuous one $[0, 1]$. We model the eligibility by a probability of being eligible. Let R^{max} be the random variable of the *rank of last offer* (see definition 4.2.1). Let r be the rank of the patient for a certain waiting-list and with a certain donor. If $R^{max} \geq r$, then the patient gets a proposal. Thus the probability we need is:

$$\mathbb{P}(R^{max} \geq r)$$

We define the rank of last offer:

Definition 4.2.1 (Rank of Last Offer)**.** When a donor becomes available, her kidneys are proposed to patients in descending score order in the waiting-list. We call rank of last offer for a donor the maximal rank to which a patient got an offer for this donor on the waiting-list. If two kidneys are proposed, we consider the maximal rank of the kidney offered last.

Remark 4.2.8. We introduced the eligibility probability through the rank of last offer, but it is also possible to use the score of last offer to define the eligibility probability, or a combination of both. However, an assumption on the distribution of the rank of last offer and/or the score of last offer should be done in practice. We will see in section 4.3.2 that the rank is convenient for this purpose and we will show a way to estimate $\mathbb{P}(R^{max} \geq r)$ in practice.

Justification of the Relaxation

This "relaxation" hides subtle mathematical arguments. We want to give an idea why this relaxation actually gives a better estimate of the Poisson parameters than the simple binary eligibility.

Let x be our patient (independent of time). Let $(T_j)_{j\geq 1}$ be the Poisson process of the times of arrival of parameter λ, $(Y_j)_{j\geq 1}$ be a sequence of i.i.d. random donors, $(W_j)_{j\geq 1}$ a sequence of random waiting-lists and $(D_{jk})_{j,k\geq 1}$ a set of independent random decisions. The index k does not refer to patients or donors, it has only a mathematical meaning: we make several i.i.d. drawings for the same donor and waiting-list and claim that the decisions of patients change each time following a certain distribution.

The decisions depend on the donor Y_j and the patients in W_j with some stochastic i.i.d variables U_{jk}: $D_{jk} = d(Y_j, W_j, U_{jk})$. The D_{jk} are i.i.d with j fixed. We assume that for each j and k: Y_j, W_j and U_{jk} are independent of each other and of T_j.

Remark 4.2.9. The U_{jk} encapsulate all the randomness of the decision which is not captured by W_j and Y_j and hides in unknown latent variables (and/or in free-will). These variables should be considered as a mathematical trick to achieve independence and identical distribution more than a philosophical consideration. Indeed, we can assume U_{jk} to be independent of Y_j and W_j, whereas this is not the case for D_{jk}.

We define $R_j = \phi(x, Y_j, W_j)$ the rank of patient x at time T_j and $R^{max}_{jk} = \chi(x, Y_j, W_j, D_{jk}) = \psi(x, Y_j, W_j, U_{jk})$ the rank of last offer. We see that the R^{max}_{jk} for j fixed are i.i.d.

To estimate the Poisson parameter of the arrival of eligible donors, we transform equation 4.1 to:

$$\bar{\lambda} = \overset{a.s.}{\lim_{t\to+\infty}} \frac{1}{t} \sum_{j=1}^{+\infty} \mathbb{1}_{T_j \leq t} \times \mathbb{P}(R_{j1}^{max} \geq R_j | R_j)$$

The fact that we are estimating a Poisson-like parameter is not obvious here. We would have a Compound Poisson process if the $\mathbb{P}(R_{j1}^{max} \geq R_j | R_j)$ were i.i.d., whereas they are only independent. Still, this would not *a priori* converge to any sort of Poisson parameter.

Before formalising the justification, we give an intuition why this is actually relevant. The probability $\mathbb{P}(R_{j1}^{max} \geq R_j | R_j)$ can be written as a random series of characteristic functions $\mathbb{1}_{R_{jk}^{max} \geq R_j}$ (because of the Law of Large Numbers). From a sum of elements in $[0, 1]$ we can get a mean over k of sum of elements in $\{0, 1\}$. And so the first sum is actually an average of Poisson parameters.

Theorem 4.2.4 (Compound-like Poisson Process as averaged Poisson Process Thinning)**.** *Let us define:* $\Lambda_k = \overset{a.s.}{\lim_{t\to+\infty}} \frac{1}{t} \sum_{j=1}^{+\infty} \mathbb{1}_{T_j \leq t} \mathbb{1}_{R_{jk}^{max} \geq R_j}$, $\forall k \in \mathbb{N}^*$, *the parameter of the Poisson process* $\sum_{j=1}^{+\infty} \mathbb{1}_{T_j \leq t} \mathbb{1}_{R_{jk}^{max} \geq R_j}$ *(thinned from the original* $\sum_{j=1}^{+\infty} \mathbb{1}_{T_j \leq t}$*). We assume that* $\mathbb{E}_{U_{j1}}(\Lambda_1)$ *exists. Then we have:*

$$\bar{\lambda} = \overset{a.s.}{\lim_{n\to+\infty}} \frac{1}{n} \sum_{k=1}^{n} \Lambda_k$$

Remark 4.2.10. Let us fix k. We said that the process $\sum_{j=1}^{+\infty} \mathbb{1}_{T_j \leq t} \mathbb{1}_{R_{jk}^{max} \geq R_j}$ was a thinning from $\sum_{j=1}^{+\infty} \mathbb{1}_{T_j \leq t}$. This is the case if the $\mathbb{1}_{R_{jk}^{max} \geq R_j}$ are i.i.d. and Bernoulli. The independence is very easy to see here, but not the Bernoulli-like underlying distribution. We have no means to show it rigorously but we may argue that this expression could be approximately rewritten as: $\mathbb{1}_{R_{jk}^{max} \geq R_j} \simeq \mathbb{1}_{Y_j \in \mathcal{D}(x)}$, as we did in theorem 4.2.1, by neglecting the influence of the waiting-list.

Proof. Let us define $N_t = \sum_{j=1}^{+\infty} \mathbb{1}_{T_j \leq t}$.

$$\begin{aligned}
\bar{\lambda} &= \overset{a.s.}{\lim_{t\to+\infty}} \frac{1}{t} \sum_{j=1}^{+\infty} \mathbb{1}_{T_j \leq t} \times \mathbb{P}(R_{j1}^{max} \geq R_j | R_j) \\
&= \overset{a.s.}{\lim_{t\to+\infty}} \frac{1}{t} \sum_{j=1}^{+\infty} \mathbb{1}_{T_j \leq t} \times \mathbb{P}(\psi(x, Y_j, W_j, U_{j1}) \geq \phi(x, Y_j, W_j) | Y_j, W_j) \\
&= \overset{a.s.}{\lim_{t\to+\infty}} \frac{1}{t} \sum_{j=1}^{+\infty} \mathbb{1}_{T_j \leq t} \times \mathbb{E}_{U_{j1}}(\mathbb{1}_{\psi(x,Y_j,W_j,U_{j1}) \geq \phi(x,Y_j,W_j)} | Y_j, W_j)
\end{aligned}$$

$$= \overset{a.s.}{\lim_{t\to+\infty}} \frac{1}{t}\sum_{j=1}^{N_t} \mathbb{E}_{U_{j1}}(\mathbb{1}_{\psi(x,Y_j,W_j,U_{j1})\geq\phi(x,Y_j,W_j)}),$$

because W_j and Y_j are independent of U_{j1}

$$= \overset{a.s.}{\lim_{t\to+\infty}} \frac{1}{t}\sum_{j=1}^{N_t} \mathbb{E}_{U_{i1},\forall i}(\mathbb{1}_{\psi(x,Y_j,W_j,U_{j1})\geq\phi(x,Y_j,W_j)}),$$

because the $U_{i1},\ W_i$ and Y_i are independent of each other

$$= \overset{a.s.}{\lim_{t\to+\infty}} \mathbb{E}_{U_{i1},\forall i}\left(\frac{1}{t}\sum_{j=1}^{N_t} \mathbb{1}_{\psi(x,Y_j,W_j,U_{j1})\geq\phi(x,Y_j,W_j)}\right),$$

because N_t is independent of U_{j1} and is finite almost surely

At this point of the proof, we would like to enter the limit sign inside the expected value. However, this cannot be directly obtained by the Dominated Convergence Theorem for random variables as the expected value is not on the random variables affected by the limit. Let $f_t(\omega,u) = \frac{1}{t}\sum_{j=1}^{N_t(\omega)} \frac{1}{t}\mathbb{1}_{\psi(x,Y_j,W_j,u_j)\geq\phi(x,Y_j,W_j)}$ (we assume Y_j and W_j to be fixed here as we condition upon those variables). In a very general framework, provided the distribution of U_{j1} is measurable, we can write:

$$\mathbb{E}_{U_{i1},\forall i}(f_t(\omega,U_{\cdot 1})) = \int f_t(\omega,u)dP,$$

with $U_{\cdot 1} = (U_{11},...,U_{i1},...)$

This equality holds almost surely, i.e. $\forall\omega\in\Omega'$, where $\overline{\Omega'} = \Omega$. Then $\forall\omega\in\Omega'$ and $\forall u$, we have:

$$|f_t(\omega,u)| \leq \sum_{j=1}^{N_t(\omega)} \frac{1}{t} = \frac{N_t(\omega)}{t} \underset{t\to+\infty}{\longrightarrow} \lambda$$

The limit holds almost surely, i.e. for almost every $\omega\in\Omega'$, but we keep this notation at the risk of redefining it. As $\frac{N_t(\omega)}{t}$ converges, it is also bounded, and $\exists M(\omega)>0$ such that $|f_t(\omega,u)| \leq \frac{N_t(\omega)}{t} \leq M(\omega)$. $M(\omega)$ is integrable: $\int M(\omega)dP = M(\omega)\int dP = M(\omega)$. Thus we can apply the deterministic version of the dominated convergence theorem $\forall\omega\in\Omega'$:

$$\lim_{t\to+\infty}\int f_t(\omega,u)dP = \int \lim_{t\to+\infty} f_t(\omega,u)dP$$

As this equality holds for almost all $\omega \in \Omega$, we can do the inversion which we wanted:

$$\bar{\lambda} = \mathbb{E}_{U_{i1},\forall i}\left(\overset{a.s.}{\lim_{t\to+\infty}} \frac{1}{t}\sum_{j=1}^{N_t} \mathbb{1}_{\psi(x,Y_j,W_j,U_{j1})\geq\phi(x,Y_j,W_j)}\right)$$

At this point, we would like to apply the Law of Large Numbers to $\Lambda_k = \overset{a.s.}{\lim_{t\to+\infty}} \frac{1}{t}\sum_{j=1}^{N_t} \mathbb{1}_{\psi(x,Y_j,W_j,U_{jk})\geq\phi(x,Y_j,W_j)}$. The Λ_k are independent of each other relatively to the U_{jk} and they are identically distributed because the U_{jk} (for j fixed) are identically distributed and they share all the other variables (as N_t, Y_j and W_j are considered deterministic under the expected value sign). We can apply the law of large numbers to the Λ_k and thus:

$$\bar{\lambda} = \overset{a.s.}{\lim_{n\to+\infty}} \frac{1}{n}\sum_{k=1}^{n} \Lambda_k$$

□

Remark 4.2.11. In the previous proof, we did not give a precise definition of all the sets over which we integrate and all the spaces for the variables, because they are very abstract. The purpose of this proof was to show how applying the Law of Large Numbers justified the relaxation, and how to proceed to the necessary sum-integral inversions.

4.2.4 Quality of Eligible Donors

Several graft quality measures exist as discussed in section 2.1. The simplest one is the age of the donor (to simplify: the younger the better). Then we can think of the KDRI (Rao et al., 2009). We can give a general form of such a quality measure. Let $q : \mathcal{P} \times \mathcal{D} \to \mathbb{R}$ be such a quality measure. We call T the time of next offer for the patient x. To refer to the patient x at time T, we write: x_T. We call Y the next eligible donor (arriving at time T by definition). Y depends on T because the distribution of eligible donors varies over time. The quality of the next offer is the random variable: $Q = q(x_T, Y)$. Let $q_0 = q(x_0, y_0)$ be the quality of the initial offer. We would like to put q_0 in perspective with Q, so the patient has an idea of the relative quality of his current offer compared to the potential future ones. Typically, the expected quality and the dispersion would be suitable.

We have to infer $\mathbb{E}(Q)$ and $\mathrm{Var}(Q)$ from our estimated distribution. The following result solves the problem:

Lemma 4.2.2 (Expected Quality)**.**

(i) $\mathbb{E}(Q) = \sum_{i=0}^{m} \mathbb{P}(t_i \leq T < t_{i+1}) \mathbb{E}(Q | t_i \leq T < t_{i+1})$

(ii) $\mathrm{Var}(Q) = \sum_{i=0}^{m} \mathbb{P}(t_i \leq T < t_{i+1}) \mathbb{E}(Q^2 | t_i \leq T < t_{i+1}) - \mathbb{E}(Q)^2$

Proof. Direct application of the law of total expectation. □

We will be able to estimate these expressions in practice, thanks to the framework we built. Indeed, we use the distribution of T that we estimate and the set of eligible donors which we find to estimate the parameters of the distribution of donors $\forall i \in \{0, ..., m\}$ between t_i and t_{i+1}.

Remark 4.2.12. These estimations are reliable as long as the dataset is sufficiently large.

This is a first approach. However, we may want to give another information to the patient. Not the expected quality for the whole distribution but the expected quality at the expected time of next offer: $\mathbb{E}(Q | t_{i_T} \leq T < t_{i_T+1})$, with i_T such that $t_{i_T} \leq \mathbb{E}(T) < t_{i_T+1})$. This can also be estimated in practice.

4.3 Algorithmic Perspective

In this section, we show how we can estimate the parameters of the distribution of T in practice (for any suitable dataset). We introduce several variants, depending on how we predict the eligibility of any donor to our patient.

4.3.1 General Algorithm

The general algorithm considers donors arrived in the past as a representative subset of the global distribution of future donors and of their arrival.

There is a risk of confusion in the notations in this section and next section: actual distributions, distributions perfectly fitted to our model, distributions estimated for our model. When we feel necessary, we will add an exponent to the concerned random variable for clarity (e.g. T: T^* (actual), $T^=$ (perfectly fitted), $T^\simeq$ (estimated)).

Input

- $(x_0, y_0) \in \mathcal{P} \times \mathcal{D}$: current offer at time $t_0 = 0$.
- ΔT: number of days from t_0 to the past to use as a "training" set of donors.

- ϵ: precision threshold on the expected value (see theorem 4.2.3).
- Several features specific to the algorithm variant.

Data

The donors $(y_j)_{j\geq 1}$ arrived from $t_0 - \Delta T$ to t_0 and the patients to which they were proposed at their time of arrival $(w_j)_{j\geq 1}$ (the past waiting-lists taken in some waiting-lists set $\mathcal{W}$).

Remark 4.3.1. This is what we use as a training set. No information about the future is needed: we make no use of the actual observed time of next offer to make the prediction as such.

Output

- $(\mu_0, ..., \mu_m)$: vector of Poisson parameters.
- $\mathbb{E}(T)$, t_α, $\mathbb{E}(Q)$... Any results which can be deduced from the distribution of T.

Algorithm

This algorithm relies on several assumptions in addition to the basic ones made for the whole work: 4.2.1, 4.2.4.

We consider an "eligibility" function $e : \mathcal{P} \times \mathcal{D} \times \mathcal{W} \to [0, 1]$ given. e will change according to the variants of the algorithm which we use. $e(x, y, w) = 0$ means ineligible, $e(x, y, w) = 1$ means eligible and any floating number in-between refers to a probability of being eligible (see section 4.2.3). Although in the real attribution process, TQ knows the anti-HLA specificities for each patient, we do not have this information in our data. As it is not possible with our data to know in advance if a patient and a donor are tissue-type compatible, this eligibility function does not account for the positive cross-match probability: eligibility means "would get an offer assuming cross-match compatibility".

The general algorithm is presented below (algorithm 1):

$\Delta\tau := +\infty$;
for t_i *while* $\Delta\tau > \epsilon$ *and* $i \leq m$ **do**
  $N := \sum_{j\geq 1} e(x_{t_i}, y_j, w_j)$; `// number of eligible donors`
  $\mu_i := N/\Delta T$;
  $\Delta\tau := e^{-\sum_{k=0}^{i} \mu_k (t_{k+1}-t_k)}/\mu_i$; `// precision on the expected value`
end
`// We take the probability of positive cross-match into account`
$\mu_\cdot := \mu_\cdot * (1 - x_0^{cPRA}/100)$

Algorithm 1: Basic pseudo-code

Remark 4.3.2. $(1 - x_0^{cPRA}/100)$ is the probability of being tissue-type compatible to any donor.

In order to get an idea of the precision of the parameters we estimate, we would like to get confidence intervals on $\mathbb{E}(T^{\simeq})$. As we mentioned earlier in section 4.2.2 we were unable to get an explicit formula for such intervals. Therefore, we will use bootstrapping to estimate them. Bootstrapping is known to generally give better intervals than simple normal assumptions (DiCiccio and Efron, 1996). For a number N_{boot} of simulations, we resample the set of donors. Let $\Phi_k : \{1, ..., n\} \to \{1, ..., n\}$ be a random resampling of the n donors in $[t_0 - \Delta T, t_0]$. We apply the algorithms to $(y_{\Phi_k(j)})_{j\geq 1}$, $\forall k \in \{1, ..., N_{boot}\}$.

4.3.2 Detailed Variants

The variants reside in the eligibility function e. We propose two main approaches, which themselves subdivide in two approaches.

We remind that e takes into account blood-type ($e(x, y, w) = 0$ if blood-type ineligible) and ranking (or score) on the waiting-list.

Past Waiting-List

We consider donor y_j, arrived somewhere in $[t_0 - \Delta T, t_0]$ on the waiting-list w_j. At that time, the rank of last offer was ρ_j. We compute the rank r_{ij} of patient x_{t_i} in w_j after removing all the blood-type ineligible candidates and cross-match positive candidates. If the patient has a better rank than the last rank ($r_{ij} \leq \rho_j$): $e(x_{t_i}, y_j, w_j) = 1$, else $e(x_{t_i}, y_j, w_j) = 0$.

Remark 4.3.3. One may wonder why we remove cross-match positive candidates from the

waiting-list, considering we claimed earlier that e did not account for cross-match compatibility. Indeed, we do not take into account cross-match compatibility of our patient. We take into account cross-match compatibility of candidates on the past waiting-list when the donor arrived and whom we know if they were compatible or not.

In practice, it is not necessary to keep track on the whole waiting-list and compute the exact rank of the patient. It is sufficient to calculate the score of the patient and to compare it with the score of last offer:

Lemma 4.3.1. *Let σ_j be the score associated with the last rank ρ_j, and $s_{ij}(= f_s(x_{t_i}, y_j))$ the score of our patient of rank r_{ij}. Then:*

$$r_{ij} \leq \rho_j \Leftrightarrow s_{ij} \geq \sigma_j$$

Proof. This is obvious because the score inverts the rank: higher score means lower rank. □

Current Waiting-List

In the previous section, we only used information from the past waiting-list to assess eligibility. However, we feel that we may miss important information by using an outdated waiting-list for our current patient. Therefore, we use the current waiting-list to better represent the actual situation of the patient and the rank of last offer from the past as a measure of popularity of the concerned donor.

We consider donor y_j, arrived somewhere in $[t_0 - \Delta T, t_0]$ in the waiting-list w_j^{past}. At that time, the rank of last offer was ρ_j. Let $w^{current}$ be the current waiting-list of the patient. We compute the rank r_{ij} of patient x_{t_i} in $w^{current}$ after removing all the blood-type ineligible candidates and cross-match positive candidates. If the patient has a better rank than the last rank ($r_{ij} \leq \rho_j$), $e(x_{t_i}, y_j, w_j) = 1$, else $e(x_{t_i}, y_j, w_j) = 0$, where w_j combines useful information from both $w^{current}$ and w_j^{past}.

There are several subtleties here:

- As t_i increases in the algorithm, should we update the age and the waiting-times of patients in $w^{current}$?
  In reality, after each t_i is updated, some patients will have stopped waiting and others will have come in. If we update the patients in the waiting-list, we assume that everybody stays in the waiting-list and has an increased waiting-time. Instead, we prefer to

assume a certain stationarity of the waiting-list: each time a patient goes out of the waiting-list, he is replaced by an identical patient with the same age as at t_0.

- How do we know if the patients on $w^{current}$ cross-match compatible to y_j?
  This time, we do not know. Therefore, we have to make simulations. We will use the cPRA of the patients on $w^{current}$ to draw at random cross-match compatible patients.

We get the following algorithm (algorithm 2):

input : n_{sim}: number of simulations
output: e: eligibility
$w^{current} :=$ filter_ABO$(w^{current}, y_j)$; `// filters blood-type ineligible patients`
`// Draws` n_{sim} `times at random cross-match compatible patients`
$Xmatch := 1. * (\text{rand}(n_{sim}) > (w^{current}_{cPRA}/100))$;
$scores = f_s(w^{current} * Xmatch)$;
$\sigma_j = scores[\rho_j, :]$; `// scores of last offer for each simulation`
$\bar{\sigma}_j =$mean(σ_j) ;
$e = 1. * (s_{ij} > \bar{\sigma}_j)$; `// eligibility`

Algorithm 2: Current Waiting-List Eligibility

Remark 4.3.4. There are several ways of reducing the information obtained through the cross-match compatibility simulations. In the above algorithm, we simply compute the mean of the scores of last offer. It would be also possible to compare s_{ij} to each of the simulated last scores and then make a majority vote: eligible if eligible in the majority of the simulations, ineligible otherwise. It would be even possible to simply return the percentage of simulations which lead to eligibility.

Eligibility Relaxation

We include the considerations of 4.2.3 in the algorithm by choosing:
$e(x_{t_i}, y_j, w_j) = \mathbb{P}(R^{max}_j \geq r_{ij})$, in any of the preceding frameworks.

A reasonable assumption for the distribution of R^{max} would be a Poisson distribution. First of all, it is approximately the distribution that we observe on the data. Furthermore it has the particular property of having expected value equal to the standard deviation. So if we expect a very low number of proposals (little maximal rank), the ranks are less stochastic: a good donor will be accepted very fast in general, but if we expect a quite high number of proposals (big maximal rank) then the influence of the single decisions is very high and the last rank can vary a lot. This gives:

Assumption 4.3.1. *The rank of last offer for the same donor under different circumstances is Poisson distributed,* $R^{max} \sim \mathcal{P}(\rho)$*:*

$$\mathbb{P}(R^{max} \geq r) = 1 - \sum_{k=0}^{r-1} \frac{\rho^k}{k!} e^{-\rho}$$

The question is now which parameter ρ to choose for the distribution. Either we choose it fixed, then we have to take for ρ what we would have taken as a maximal rank, or we can use the historical maximal rank of acceptance for the donor we consider at the time she was released.

Quality Analysis

In section 4.2.4, we explained how to gather information from each interval $[t_i, t_{i+1}]$ to get a complete quality estimation. However, we did not explain how to estimate
$\mathbb{E}(Q|t_i \leq T \leq t_{i+1})$.

At each time step t_i of the algorithm, we have a set of eligible donors whose quality $(q_j)_{j\geq 1}$ we know. We just have to average them to get an estimate of the mean quality at this time step:

$$\mathbb{E}(Q|t_i \leq T \leq t_{i+1}) = \frac{\sum_j e(x_{t_i}, y_j, w_j) q_j}{\sum_j e(x_{t_i}, y_j, w_j)} \tag{4.2}$$

Time to Next Better Offer

It is possible to put a constraint on the quality to estimate the distribution of time to next better offer. Let q_0 the minimal acceptable quality for the patient (e.g. the quality of the current offer). We remove from the training set all the donors with a lower quality ($Q < q_0$) and we run the algorithm on the new training set.

4.3.3 Complexity Analysis

The different variants of the algorithms which we presented previously have different computational complexities. They depend on the fact that we have to compute the scores, the ranks or make simulations.

In all these analyses, we assume that the computation of the σ_j and the ρ_j was made in preprocessing. We also assume that the waiting-list at t_0 is given and there is no need to retrieve it from a general list of patients at all times (which can be the case in practice when working on retrospective data).

m is the number of μ_i which we compute. n is the number of donors which we have between $t_0 - \Delta T$ and t_0. p is the number of patients in the waiting-list at time t_0. n_{sim} is the number of cross-match simulations.

Theorem 4.3.1 (Computational Complexity)**.** *Complexity analysis of the algorithms to estimate the parameters of the non-homogeneous Poisson process.*

(i) Past Waiting-List: $\mathcal{O}(m(n+m))$

(ii) Current Waiting-List: $\mathcal{O}(n_{sim} np \ln(p) + m(n+m))$

Remark 4.3.5. We do not take into account bootstrapping in this analysis. We do not take into account the case of probabilistic eligibility. In practice, the complexity gets more complicated, because of the fact that we are using retrospective data, for which we always have to retrieve the donors in a certain time-span and the patients active on the waiting-list at t_0. Moreover, for the Current Waiting-List method, we assumed that a mean last score $\bar{\sigma}_j$ was computed for each donor, and not a "majority vote" on the eligibility of each donor, which also changes the complexity. Finally, the filtering of the blood-type ineligible donors beforehand can also reduce complexity, as well as the early-stopping criterion (section 4.2.3).

Proof.

(i) Past Waiting-List: the first factor m comes from the `for` loop. The first term of the second factor n comes from the n comparisons between s_{ij} and σ_j. The second term of the second factor m comes from the stopping criterion.

(ii) Current Waiting-List:
$\mathcal{O}(n_{sim} np \ln(p))$: as we assume the waiting-list to be stationary over the t_i, we can compute the minimal score of acceptance once and for all before the main `for` loop. For each cross-match simulation, we have to sort the waiting-list ($p \ln(p)$) on all the n donors.
$\mathcal{O}(m(n+m))$: same explanation as for the past waiting-list algorithm.

□

4.4 Summary

- The arrival of donors in the waiting-list is following a Poisson point process with constant arrival rate (assumption 4.1.5).
- For a certain offer (x, y), we consider all the donors who arrived up to a time ΔT before the current time t_0. See figure 4.1 for an illustration.
- For each donor y_j, we compare the historical rank of last offer R_j^{max} to the rank r_j of the patient x_{t_0} for such a donor. To compute the rank of the patient for the historical donors, we can use either the historical waiting-list or the waiting-list of the patient at time t_0 (i.e. the current waiting-list).
- We estimate the arrival rate of eligible donors μ_0 by averaging the number of eligible donors over ΔT (equation 4.1): $\mu_i = \frac{\sum_{j\geq 1} \mathbb{1}_{R_j^{max}\geq r_j}}{\Delta T}$. Alternately, we can do a so-called relaxation. We compute the probability of eligibility of each donor $\mathbb{P}(R_j^{max} \geq r_j)$ and average the probabilities over ΔT (theorem 4.2.4): $\mu_i = \frac{\sum_{j\geq 1} \mathbb{P}(R_j^{max} \geq r_j)}{\Delta T}$.
- At each waiting-anniversary t_i, the waiting-score of the patient changes (assumption 4.2.4), and we take into account the age and waiting-time change by updating the arrival rate μ_i (theorem 4.2.2 and equation 4.1).
- We repeat the same process as before for as many μ_i as needed until we have a good estimation of the distribution (theorem 4.2.3).
- Once the distribution is estimated through the μ_i, we can infer all relevant information: expected time to next offer, time to next offer with 95% confidence, expected quality of next offer, etc. At no point of the process, we used information posterior to t_0.

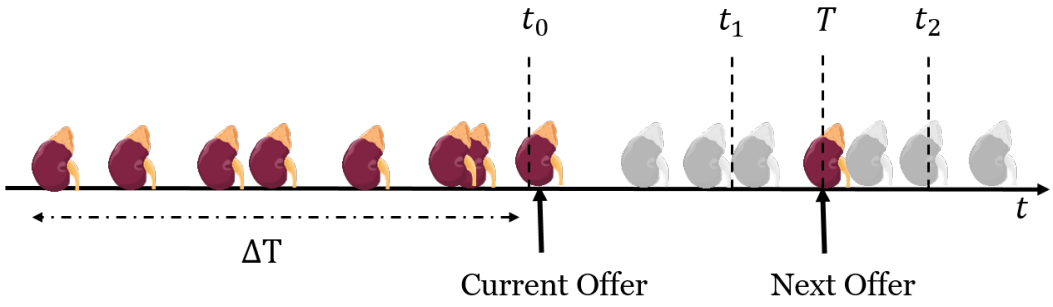

Figure 4.1 Illustration of the algorithm to estimate the distribution of time to next offer.

CHAPTER 5 OBSERVATIONAL DATA

After developing our theoretical model, we then set out to test it using data provided by TQ between 2012-03-29 and 2017-12-13 (included). We had the chance to be in contact with the persons in charge of data management at TQ so we could fully understand the dataset.

Throughout this work, we used Python 2.7.13.

5.1 Preprocessing

5.1.1 Files and Features

The data was provided in three different `.csv` files:

1. A `donor` file: containing multiple lines for each donor, each of them relating to a kidney offer with corresponding patient ID.
2. A `patient` file: containing multiple lines for each patient, each of them relating to a change in the patient's status: permanently or temporarily removed from waiting-list, transplanted.
3. A `patient_cpra` file: containing multiple lines for each patient, one per cPRA update. Indeed, the cPRA is measured regularly, to capture potential variations due to exposure to antigens (transplant, transfusion, pregnancy).

For each of these files, we had different features. We enumerate the features we kept (basically all the standardised non hand-written features) in tables 5.1, 5.2, 5.3.

Remark 5.1.1. The fact that we kept these features does not mean we used them all. We kept some features in the perspective of potentially using them later on. For instance, in case the algorithm we use to predict the quality of a transplant requires it.

5.1.2 Formatting

In tables 5.1, 5.2, 5.3, we introduced the features after formatting. We converted all categorical string features (excepted for blood-type which we let in string) in integers. We standardised the dates in the format `year-month-day` and removed the time of the day as

Table 5.1 Features of the donor file.

Feature	Signification	Type
DON_ID	Donor Identification Number	Integer
DON_BTH_DT	Donor Birth Date	YYYY-MM-DD
DON_STATUS	Donor Status (e.g. DND or DCD)	Integer
DON_DEATH_TM	Donor Date of Death	YYYY-MM-DD
DON_AGE	Donor Age in years	Integer
DON_GENDER	Donor Gender (Male, Female)	1, 2
DON_DIAB	Donor History of Diabetes (Never, Former, Active)	0, 1, 2
DON_COCAINE	Donor History of Cocaine (Never, Former, Active)	0, 1, 2
DON_CIGARETTE	Donor History of Cigarette (Never, Former, Active)	0, 1, 2
DON_CORONARY	Donor Coronary Disease	0, 1
DON_VASC	Donor Vascular Disease	0, 1
DON_HTN	Donor History of Hypertension (Never, Former, Active)	0, 1, 2
DON_CREAT	Donor Creatinine (μmol.L^{-1})	Integer
DON_WGT_KG	Donor Weight in kg	Integer
DON_HGT_CM	Donor Height in cm	Integer
DON_RACE	Donor Race	Integer
DON_COD	Donor Cause of Death	Integer
DON_EXC	Donor Exceptional Distribution	0, 1
DON_ABO	Donor Blood-Type	Character
DON_RH	Donor Rhesus	0, 1
DON_ANTI_HCV	Donor Hepatitis C Serology	Integer
DON_A/B/BW/CW /DQ/DR/DRW_1/2	Donor HLA Allele A, B, BW, CW, DQ, DR or DRW at Locus 1 or 2	Integer
DON_ORG	None, Left or Right Kidney	0, 1, 2
DON_RCV	Donor Kidney Recovered	0, 1
DON_WHY_NOT_RCV	Donor Why Not Recovered	Integer
DON_WHY_REFUSED	Donor Why Refused	Integer
DON_WHY_FAMILY _REFUSED	Donor Why Family Refused	Integer
DON_WHY_NOT_TX	Donor Why Not Transplanted	Integer
DON_WHAT_IF_NO _TX	What happened to the organ if not transplanted	Integer
CAN_LIST	Priority list of the Candidate offered the kidney	Integer
DON_CAN_SCORE	Donor Candidate Score if relevant	Float
CAN_RANK	Rank of the Candidate	Integer
CAN_DECISION	Rejection or Acceptance	0, 1
CAN_STATUS	Candidate Status (transplanted or not)	Null, 2
CAN_WHY_NO	Why Candidate Refused	Integer
CTR_NO_FOR_ALL	Centre Refused for all candidates	0, 1
CAN_CTR_ID	Candidate Centre Identification Number	Integer

Table 5.2 Features of the `patient` file.

Feature	Signification	Type
CAN_ID	Candidate Identification Number	Integer
CAN_BTH_DT	Candidate Birth Date	YYYY-MM-DD
ORG_TY	Organ to which the Candidate is applying	Integer
CAN_GENDER	Candidate Gender (Male, Female)	1, 2
CAN_WGT_KG	Candidate Weight in kg	Integer
CAN_HGT_CM	Candidate Height in cm	Integer
CAN_ABO	Candidate Blood-Type	Character
CAN_RH	Candidate Rhesus	0, 1
CAN_AGHBS	Candidate Hepatitis B Serology	Integer
CAN_ANTI_HCV	Candidate Hepatitis C Serology	Integer
CAN_ANTI_HIV	Candidate HIV Serology	Integer
CAN_A/B/BW/CW /DQ/DR/DRW_1/2	Donor HLA Allele A, B, BW, CW, DQ, DR or DRW at Locus 1 or 2	Integer
CAN_CPRA	Candidate latest cPRA	Integer
CAN_CPRA_DT_TM	Latest Date of cPRA measurement	YYYY-MM-DD
CAN_LISTING_DT	Candidate latest Date of Enlisting	YYYY-MM-DD
CAN_DIAL_DT	Candidate latest Date of First Dialysis	YYYY-MM-DD
CAN_WTG_DT	Starting Date for the waiting Chronometer	YYYY-MM-DD
CAN_NB_TX	Number of Transplant the Patient underwent	Integer
CAN_STATUS	Candidate Status on Waiting-List (deceased, inactive, active, DDKT)	-1, 0, 1, 2
UPDATE_TM	Date of Status update	YYYY-MM-DD
CAN_WHY_RMV	Why the Candidate was removed from the waiting-list if relevant	Integer
CAN_DGN	Initial Diagnosis	Integer
CAN_DGN2	Secondary Diagnosis	Integer

Table 5.3 Features of the `patient cpra` file.

Feature	Signification	Type
CAN_ID	Candidate Identification Number	Integer
CAN_CPRA	Candidate cPRA	Integer
UPDATE_TM	Date of cPRA update	YYYY-MM-DD

we are only need a day precision. This should not be problematic in terms of ties as the equation 4.1 we use to estimate the parameters only uses the number of arrivals between two dates.

We introduced the feature *waiting-date* of a patient (`CAN_WTG_DT`) in table 5.2. We created this feature from the date of first dialysis and the date of enlisting according to the rule given in section 3.1.2. When the date of first dialysis was missing, we assumed that the patient was not under dialysis, and thus would not get waiting points in the new policy.

5.1.3 Missing Data

We were in direct contact with TQ for the data. When data was missing, it was not because these data did not exist, but often because they were gathered elsewhere.

- HLA, cPRA, weight and height values were missing for some patients. They were collected from another source and conveyed to us.
- Weight, height, cause of death, gender, hypertension and creatinine were missing for donors between 2012 and 2015 roughly. They were collected from paper records as well as additional features (useful for the prediction of graft quality): history of diabetes, cocaine, cigarette, coronary disease, vascular disease.

5.1.4 Cleaning

As in any empirical dataset, there are errors. These errors could originate at three different steps:

- When the information was hand-entered in the database.
- When the information was extracted from the database to a `.csv` format.
- When missing or new relevant features were compiled from paper records (cPRA, HLA, gender, history of diabetes, cocaine, cigarette, hypertension, coronary disease, vascular disease, weight or height).

We tried to identify basic inconsistencies and mistakes in the data.

- We removed one line of a donor with an invalid string ID (this donor had not been proposed to anyone on the list)

- We corrected by hand some obvious mistakes for a dozen of donors (e.g. ID 176 next to ID 2672 donors with same characteristics would be renamed 2672)

- We calculated the age from the date of birth of the donor at time of death and compared it to the `DON_AGE` feature. For 5 donors, we had a minor (1 year) difference.

- We removed 253 rows (corresponding to 79 patients) out of 14293 rows (corresponding to 3665 patients) with dates mismatches: update time anterior to enlisting date and date of first dialysis.
  This is an annoying mistake due in majority to retransplant problems (see 3.1.2). In the data, we only have the latest first dialysis and enlisting dates. When a candidate has a first transplant that functions for more than 3 months and then experiences graft failure, the patient can be registered again on the TQ waiting-list. However, because the graft survived for over 3 months, the date of the first dialysis session or waiting-list registration (which is used to compute waiting-time) is updated as the date of new dialysis onset after graft loss. For some patients, this update was missing. For other patients, the date was updated but TQ still used the initial date of dialysis onset or waiting-list registration to compute waiting-time for unknown reasons. Removing them is unsatisfying because we alter the actual waiting-list at that time, but keeping them is also problematic as we cannot compute their score correctly. Considering that only 2% of patients are concerned, we prefer removing the rows, as removing a patient with high scores should have less impact on the waiting-list than having a patient with unduly high scores. Little scores are not problematic because there are lots of low-scoring patients.

In order to identify other potential errors and to ensure our understanding of TQ's scoring function, we verified our computation of the score for each suitable offer to the value given by TQ in the column `DON_CAN_SCORE`. We only kept patients in the general attribution list. In the very end, we get less than 2% of errors: 225 mismatches for 15634 scores.

- 32 mismatches are less than or equal to 0.05 points. Such mismatches are often due to mistakes in the younger patient priority (age score). We were unable to correct these mistakes.

- The remaining 193 mismatches are greater than or equal to 0.5 points (no mistakes between 0.05 and 0.49), with a maximum of 12 for one occurrence. When they occur, they occur most of the time for all offers of a patient. From this point, we can only make assumptions. When the errors are of 3 points, it is either due to an error in

the HLA matching or cPRA matching (so we have either a wrong cPRA, less than 20 instead of between 20 and 80, or a wrong HLA type, with homozygous alleles instead of heterozygous...). When the errors are different, they most likely originate from an error in waiting-time: error in the date of first dialysis or/and enlisting. This corroborates our previous statement about retransplant and dates mismatches.

5.2 Main Figures

From this dataset, we can portray patients and donors in the province of Québec. We give information about the attribution process and parameters necessary for the attribution process (blood-type, age, patient cPRA) as well as additional parameters describing patients and donors (sex, donor race, donor BMI, patient waiting-time from enlisting.

5.2.1 Attribution in the province of Québec

We would like to quantify the proportion of kidneys which are discarded after recovery and the weight of the general attribution list in the whole process. Figure 5.1 shows the number of kidneys involved in the whole attribution process in the dataset. Ninety percent of recovered kidneys are transplanted and 90% of transplanted kidneys are attributed in the general attribution scoring list.

This shows that our work focuses on the vast majority of deceased donor transplantations in the province of Québec. Therefore, it can impact the majority of kidney transplant candidates.

5.2.2 Donors in the province of Québec

All the statistics which we provide are for donors whose kidneys were actually recovered. However, we do not wish to provide statistics about the rate of discarded donors. Indeed, there are multiple reasons why a donor is discarded, either because of her family or her health state or the condition of the organs after surgery.

We present in table 5.4 statistics about categorical features. On the 848 donors observed in the study, almost 95% are Caucasian, with a light majority of men. The distribution of blood-types matches overall Canadian population (Canadian Blood Services, 2018). We also give statistics on several numerical features in table 5.5, which show that the age distribution is quite concentrated around the median. We also provide information about the distribution of

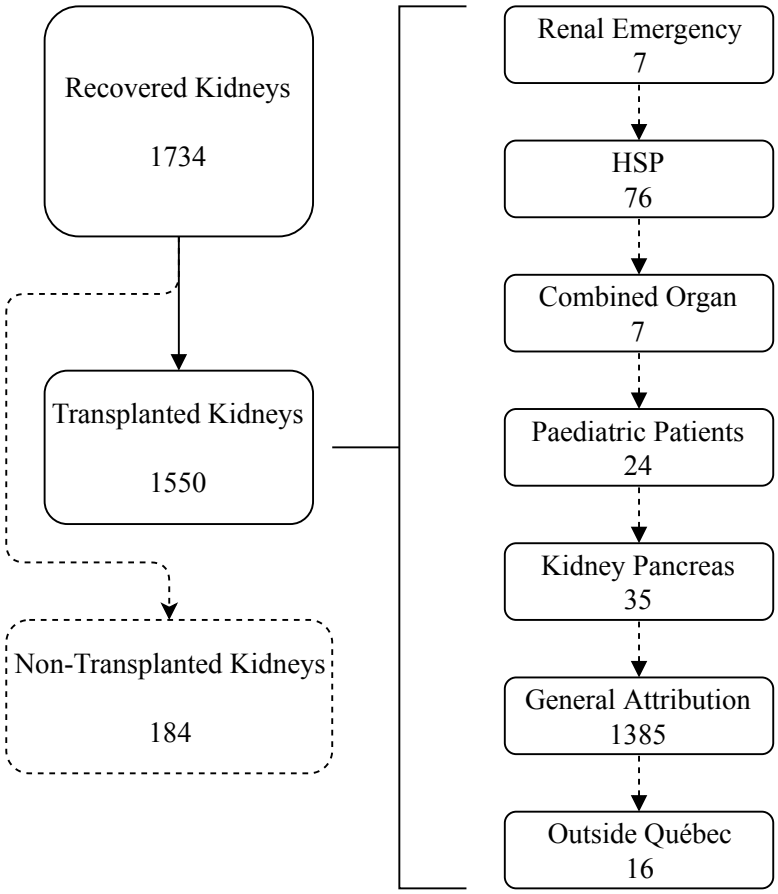

Figure 5.1 Numbers of kidneys recovered, transplanted and attributed in the different priority lists in Québec between 2012-03-29 and 2017-12-13.

rank of last offer and the corresponding minimum score. We neglected two infinite minimum scores, corresponding to the fact that the kidneys did not make it to the scoring waiting-list and were accepted in the priority lists. We see that the minimum scores, last rank and number of offers outside the waiting-list are very concentrated around the median, with extreme values though.

Remark 5.2.1. The last rank of offer is the rank of the lowest-scoring patient on the waiting-list who actually got an offer. We do not consider the rank of transplantation, because some of the organs were not transplanted or only one organ was transplanted and the other was not. Moreover, we are interested in the fact that patients got an offer for donor eligibility, not in the fact that this offer resulted in a transplantation.

We provide a plot of the distribution of the ranks of last offer. We remind that this is the rank on the scoring waiting-list of the patient who got the last offer after the kidney was

Table 5.4 Distribution of categorical features of recovered donors in Québec between 2012-03-29 and 2017-12-13.

Feature	Category	Number	%
Race	Arab	7	0.8
Race	Asian	13	1.5
Race	Black	7	0.8
Race	Caucasian	801	94.5
Race	Latin America	4	0.5
Race	Native	8	0.9
Race	Other	5	0.6
Race	Unknown	3	0.4
Status	DCD	116	13.7
Status	DND	732	86.3
Blood-Type	A	357	42.1
Blood-Type	AB	32	3.8
Blood-Type	B	85	10.0
Blood-Type	O	374	44.1
Gender	Female	372	43.9
Gender	Male	476	56.1

Table 5.5 Distribution of numerical features of recovered donors in Québec between 2012-03-29 and 2017-12-13. "Std" refers to the standard deviation and the "x%" to the quartiles.

Feature	Count	Mean	Std	Min	25%	50 %	75%	Max
Offers outside the scoring waitlist	848.0	0.99	1.04	0.0	0.0	1.0	2.0	6.0
Max Rank of Offer	848.0	4.78	5.97	0.0	2.0	3.0	5.0	81.0
Min Score of Offer	846.0	10.15	4.11	0.67	7.8	9.8	12.08	31.1
Age	848.0	50.97	17.14	2.0	41.0	54.0	63.25	85.0
BMI	848.0	26.91	5.63	1.98	23.31	26.12	29.71	58.29

either discarded or transplanted. We fitted a Poisson distribution to the empirical distribution through maximum-likelihood and to the histogram through least-square method to put in perspective assumption 4.3.1. This assumption is about the distribution of maximal ranks of offer for the same donor, and not for different donors. Thus, it only gives an illustration of the distribution if all donors were considered of equivalent popularity.

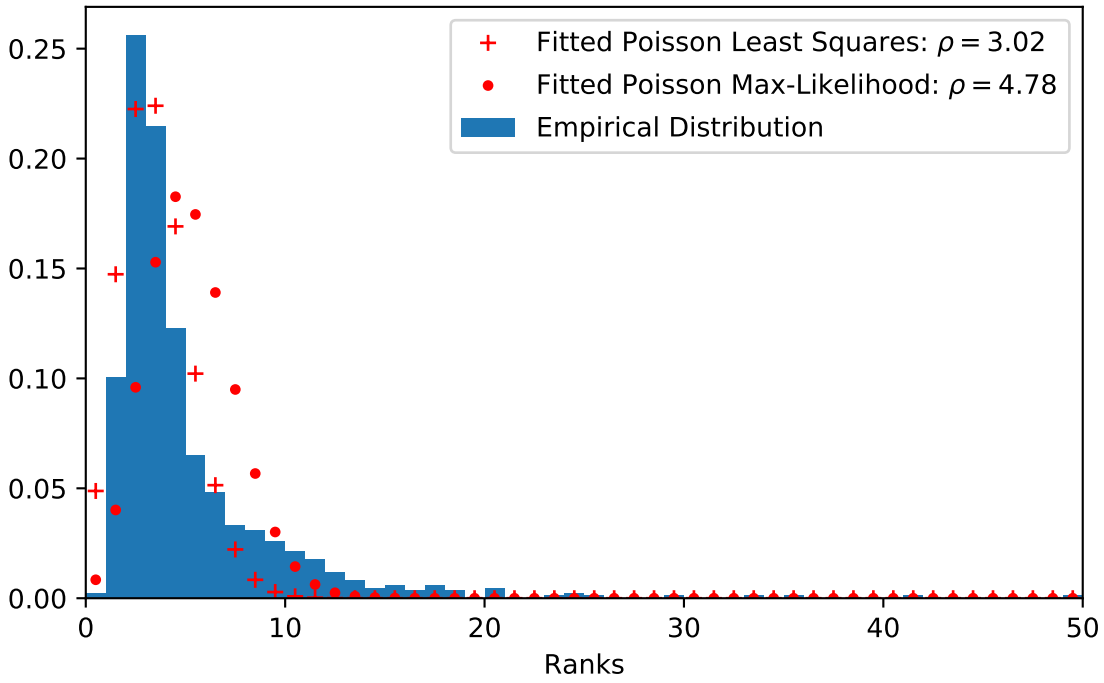

Figure 5.2 Distribution of the maximal ranks of offer for recovered donors in Québec between 2012-03-29 and 2017-12-13.

5.2.3 Patients in the province of Québec

There are several possibilities to portray the transplant candidates, as their features depend on the time of focus. We chose to present two pictures of the patients in Québec: one at the time of enlisting, one at the end of the study (including only competing patients on the general attribution list). At time of enlisting, we considered all the patients enlisted between 2012-03-29 and 2017-12-13. At the end of the study, we considered all the patients active on the general waiting-list on 2017-12-13. We included the priority candidates excepted combined organ candidates as they may fall into the general waiting-list. We also included the patients with time of enlisting anterior to the beginning of the study when they were still

active.

It is a challenge to retrieve the waiting-list at a certain time from our data. Therefore, some characteristics are missing whereas they are essential to the computation of a score, because we included in the waiting-list some priority patients who might never get into the general one. This is the reason why some cPRAs are missing. Some weights and heights are missing too, but we simply neglected them.

Some categorical features are given in tables 5.6 and 5.7. The first observation that we can make on the blood-type distribution is that it matches the overall Canadian distribution at time of enlisting but changes when censored at a later time. The proportion of A and AB patients decreased while the proportion of O and B patients increased. This may be partly due to access to living donation: it will be more difficult for O patients to find a compatible living donor than for A, B or AB patients. Thus, the proportion of O patients is greater on the waiting-list. The gender does not vary a lot, with always a majority of men among patients. This majority is slightly more important than for donors. However, the cPRA distribution shows some changes. As we expect, there is a majority of low cPRA patients which decreases a little between enlisting and time of censoring. Yet, the proportion of highly sensitised patients increases greatly while the proportion of moderately sensitised patients decreases. On the one hand, highly sensitised patients are supposed to be compatible with less donors, so it is logical they get less offers and remain more time on the waiting-list. On the other hand, they are favoured by TQ's scoring function (see section 3.2.1), as a fairness criterion. This shows that the scoring system does not totally rebalance the access to kidney with cPRA but tends to favour moderately sensitised patients over highly sensitised patients.

We give some numerical features in tables 5.8 and 5.9. As expected, the BMI distribution remains roughly identical: weight and height are not used in the score and, all other things being equal, should remain the same. More surprisingly, the age distributions are also very similar. This is more surprising, for the scoring function favours young patients over old ones. In our opinion, it is not due to a change in the distribution of patients (this supported by the comparison between age distributions in table 5.11 and 5.9, where age distribution does not change either). This seems to mean that the young patient score is of less importance than other parameters. The careful reader might be surprised to find paediatric patients in those statistics as our study focuses on non-paediatric patients. Paediatric patients were included in the set of patients competing on the scoring waiting-list, but we did not make predictions on them.

Table 5.6 Distribution of categorical features of patients at time of enlisting in Québec between 2012-03-29 and 2017-12-13.

Feature	Category	Number	%
Blood-Type	A	637	37.6
Blood-Type	AB	84	5.0
Blood-Type	B	232	13.7
Blood-Type	O	743	43.8
Gender	Female	620	36.6
Gender	Male	1076	63.4
Organ Type	Double Kidneys	1	0.1
Organ Type	Heart-Kidney	9	0.5
Organ Type	Kidney	1621	95.6
Organ Type	Liver-Kidney	18	1.1
Organ Type	Unknown	47	2.8
cPRA	Unknown	26	1.5
cPRA	[0; 20]	1091	64.3
cPRA	]20; 80]	369	21.8
cPRA	]80; 100]	210	12.4

Table 5.7 Distribution of categorical features of patients competing on scoring waiting-list on 2017-12-13 in Québec.

Feature	Category	Number	%
Blood-Type	A	66	18.5
Blood-Type	AB	9	2.5
Blood-Type	B	80	22.5
Blood-Type	O	201	56.5
Gender	Female	141	39.6
Gender	Male	215	60.4
Organ Type	Kidney	356	100.0
cPRA	Unknown	4	1.1
cPRA	[0; 20]	211	59.3
cPRA	]20; 80]	38	10.7
cPRA	]80; 100]	103	28.9

In table 5.9, we also give statistics about the waiting-time from time of enlisting. We will provide more detail on the waiting-time and the evolution of the waiting-list in the following section.

5.2.4 Evolution of the Waiting-List

Additional Figures

In previous section 5.2.3, we gave a broad comparison of the distribution of patients at their time of enlisting and the distribution of patients in the retrieved waiting-list on 2017-12-13. We give below the same figures on 2012-03-29 in tables 5.10 and 5.11. We want to give an idea of the qualitative evolution of the waiting-list from the change in attribution policy to five years later.

We observe no change in BMI, age or gender distribution between 2012 and 2017. There is no major change in highly sensitised (cPRA $> 80\%$) patients, but there was a transfer from moderately sensitised patients to non-sensitised patients between 2012 and 2017. This confirms the observation of the preceding section, that the new attribution policy favours moderately sensitised patients. However, we cannot explain very well the changes in blood-type distribution. The difference between the distribution of patients' blood-types between the overall patients' population and the patients on waiting-list increased between 2012 and 2017. This seems paradoxical because the allocation policy of TQ does not favour any blood-type (excepted for cPRA $> 1\%$ AB patients who can benefit from A and B donors). This may be caused by the unequal access to living donation: O patients can only receive kidneys from O living donors, whereas AB patients can receive kidneys from all blood-types living donors. Another explanation resides in the other priority lists: only blood-type compatibility is needed in the priority lists to get a transplant. As O donors can benefit to all priority patients, O patients get disadvantaged and stay longer in the waiting-list. The waiting-time decreased with the new allocation system, but there is still a great dispersion of the waiting-times among patients.

Table 5.8 Distribution of numerical features of patients at time of enlisting in Québec between 2012-03-29 and 2017-12-13.

Feature	Count	Mean	Std	Min	25%	50 %	75%	Max
BMI	1668.0	26.86	5.04	12.7	23.23	26.57	30.28	49.47
Age	1696.0	51.67	14.65	1.0	43.0	54.0	63.0	84.0

Table 5.9 Distribution of numerical features of patients competing on scoring waiting-list on 2017-12-13 in Québec.

Feature	Count	Mean	Std	Min	25%	50 %	75%	Max
BMI	343.0	27.14	5.12	15.47	23.56	26.85	30.4	50.35
Age	356.0	52.79	13.68	7.0	43.0	54.0	63.0	80.0
Waiting-Time (days)	356.0	996.03	1436.04	0.0	167.25	416.5	1021.0	9892.0

Table 5.10 Distribution of categorical features of patients competing on scoring waiting-list on 2012-03-29 in Québec.

Feature	Category	Number	%
Blood-Type	A	257	35.5
Blood-Type	AB	12	1.7
Blood-Type	B	93	12.8
Blood-Type	O	362	50.0
Gender	Female	278	38.4
Gender	Male	446	61.6
Organ Type	Kidney	724	100.0
cPRA	Unknown	4	0.6
cPRA	[0; 20]	349	48.2
cPRA	]20; 80]	153	21.1
cPRA	]80; 100]	218	30.1

Table 5.11 Some statistics about numerical features of patients competing on scoring waiting-list on 2012-03-29 in Québec.

Feature	Count	Mean	Std	Min	25%	50 %	75%	Max
BMI	652.0	26.43	5.12	12.52	22.7	25.9	29.54	50.35
Age	724.0	51.97	13.69	4.0	42.0	53.0	63.0	80.0
Waiting-Time (days)	718.0	1058.73	1177.97	0.0	296.25	728.0	1483.0	12237.0

Waiting-Time

We are interested in the median waiting-times from enlisting or first dialysis and to transplantation throughout the study. What we presented so far was punctual in time: 2012-03-29 and 2017-12-13. This, however, only captures the waiting-time of patients active on the waiting-list. It does not account for transplanted patients and for censored patients, i.e. patients who died or were permanently removed from the waiting-list. Similarly, we would like to know how often patients die on the waiting-list, while accounting for censored patients (transplanted or removed from waiting-list).

To this purpose, we use the well-known Kaplan-Meier estimator and draw Kaplan-Meier survival curves (Kaplan and Meier, 1958). Kaplan-Meier estimators give the probability of an event against time in a context of randomly censored data. The curves are given in figure 5.3. We removed incoherent waiting-times: 1 negative waiting-time from first dialysis and 6 negative waiting-times from enlisting. The median waiting-time for a transplant from enlisting is 1.2 years and 3.5 years from the first dialysis. For the death on waiting-list, we cannot estimate the median time of death for waiting transplant candidates. The number of deaths is too small to make the Kaplan-Meier estimator reliable, and the duration of the study is too short, in comparison to what we would expect of a survival time under dialysis. Furthermore, we may introduce a bias in our censored observations, as we consider as censored all patients who are removed from the waiting-list. Nonetheless, most of these patients remained under dialysis and we lost track of them, whereas they are very likely to have died little time after stepping out of the waiting-list (e.g. if they were removed because of worsening of their health condition). Therefore we also present the time to death or permanent removal. The median time of death or permanent removal from the first dialysis is also infinite, but the median time of death or permanent removal from enlisting is 4.7 years. This means that after 5 years waiting for a kidney transplant in the waiting-list without getting one, 50% of the population is either permanently removed or deceased. This is positive, because it means that the median time to transplant from enlisting is 4 times lower than the time to either death or permanent removal.

Evolution in Size

We provide in figure 5.4 the evolution of the size of the waiting-list throughout the study. The definition of the size of the waiting-list relies on the explanations made in section 5.2.3. We see indeed an important decrease in the size of the waiting-list between 2012 and 2017. This is consistent with CIHI (2017), as we explained in the introduction 1.2. However, it is

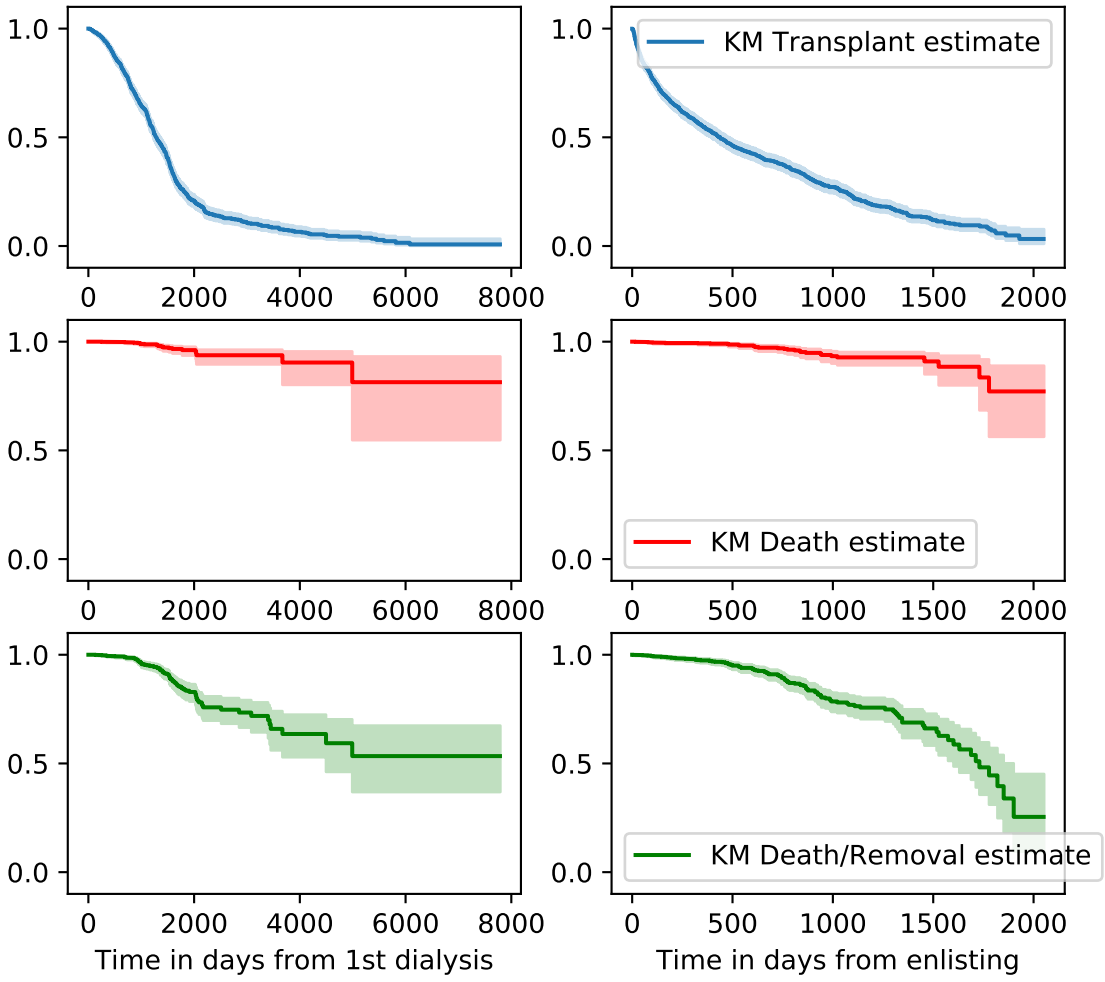

Figure 5.3 Kaplan-Meier curves with confidence intervals presenting time to transplant, time to death and time to death or permanent removal from first time of dialysis and enlisting. Patients enlisted between 2012-03-29 and 2017-12-13.

not in our capacity to disentangle what is due to the increase in deceased donation rate and what is due to the change of policy of TQ.

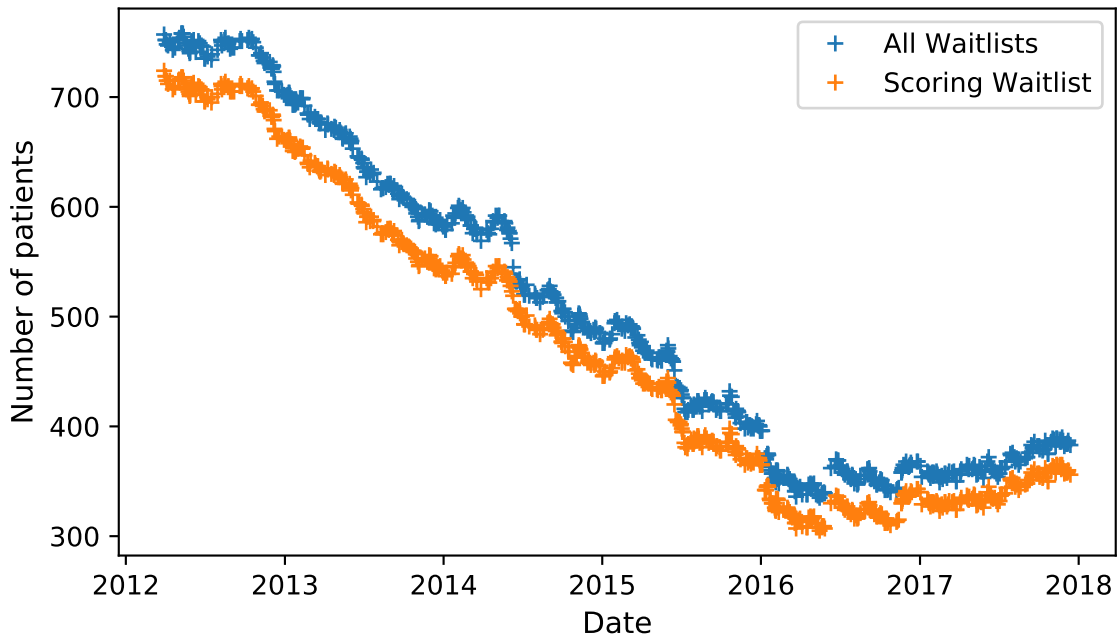

Figure 5.4 Evolution of the size of the waiting-list between 2012-03-29 and 2017-12-13. With and without multiple organs candidates.

5.3 First Verifications

We made a strong assumption 4.1.5 in the previous chapter. Before even implementing our algorithm to predict the next kidney offer, we first verify if this assumption is reasonable, i.e. if the donors arrive following a Poisson process.

We provide in figure 5.5 the value of the Poisson parameters with confidence intervals estimated with assumption **??**, assuming the arrival is Poisson. We can see that, even though the parameter does not change dramatically with time, there seems to be an increase. This is consistent with the trend we highlighted in section 1.2. The mean inter-arrival time is between 2 and 3 days (with confidence intervals) for all these parameters, under assumption ??.

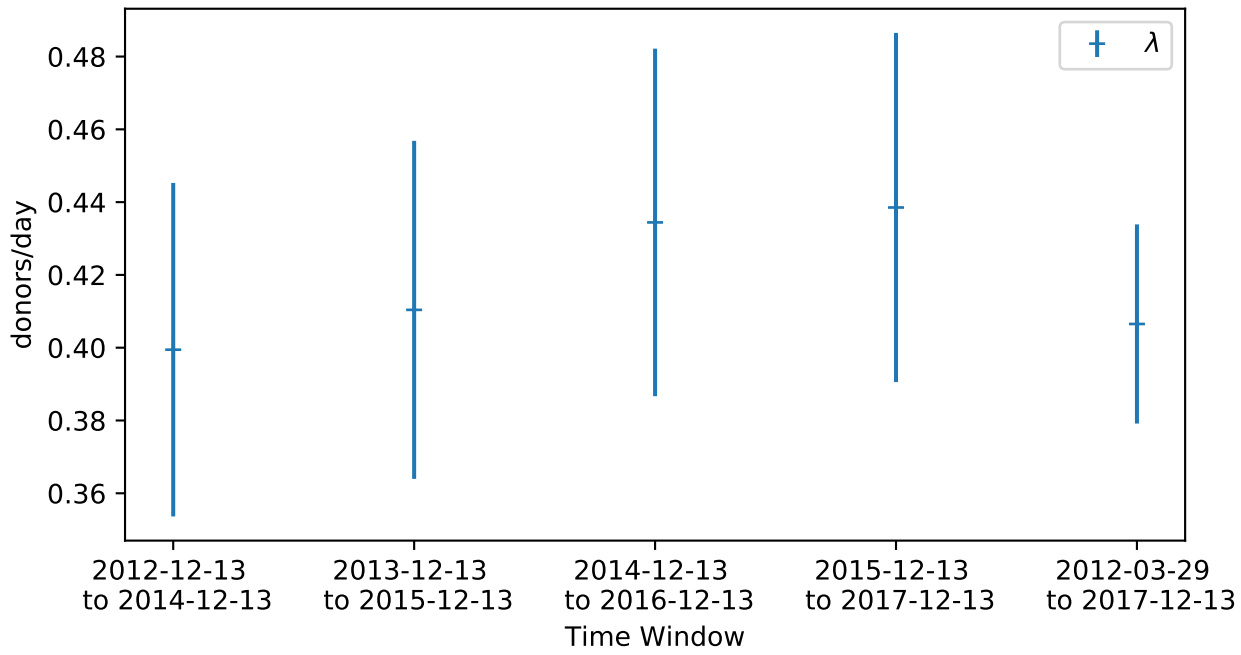

Figure 5.5 Estimation of the value of the Poisson parameter for different time windows.

Thus, our intuition is that the Poisson assumption is probably reasonable on smaller time windows and not over the whole study, considering the slow increase we observe.

In order to confirm this intuition, we can deploy several strategies. Either statistical tests or more visual verifications.

All these strategies rely on the following result which we quote from Law (2015) p. 358. Let λ be the parameter of a Poisson process $(\hat{T}_i)_{i\geq 1}$. We assume that we observed n events in the interval $[0, t]$. Let Z_1, ..., Z_n be n independent and uniformly distributed random variables on $[0, t]$. Let $Z_{(1)}$, ..., $Z_{(n)}$ be n random variables, reordered from Z_1, ..., Z_n, so that: $Z_{(1)} < ... < Z_{(n)}$ (we will not give a rigorous definition of this process here). Then $(T_1, ..., T_n)$ and $(Z_{(1)}, ..., Z_{(n)})$ have same joint distribution. This means intuitively that if one looks at an observation of n times of arrival from a Poisson process in $[0, t]$, without considering the order of arrival, she might think that n uniformly distributed random variables have been drawn. It means even that these two ways of seeing are equivalent.

Let $\hat{t}_1 < ... < \hat{t}_n$ n observed times from the process which we want to know if it is Poisson or not.

5.3.1 Kolmogorov-Smirnov Test

Background

We want to know the confidence with which we could reject the null hypothesis 4.1.5. We hope that we will not be able to reject it with a high confidence, to support the null hypothesis. To this purpose, we will compare the theoretical CDF $F(t) = t$ to the empirical CDF $F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\hat{t}_i \leq t}$ (and $F_n(\hat{t}_i) = \frac{i}{n}$). Let:

$$D_n = \sup_{t' \leq t} |F_n(t') - F(t')| = \max\left(\max_{i \leq n} |F_n(\hat{t}_i) - F(\hat{t}_i)|, \max_{i \leq n} |F_n(\hat{t}_{i-1}) - F(\hat{t}_i)|\right)$$

We use the Kolmogorov-Smirnov test applied to the Poisson point process from Law (2015) p. 358. This is the simplest case where all parameters of the theoretical distribution F are known. We have to compare: $\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right) D_n$ to a critical value $c_{1-\alpha}$, where α is the probability that we reject wrongly the null hypothesis. We reject with $(1-\alpha)$ confidence if $\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right) D_n > c_{1-\alpha}$. The order of magnitude for these critical values is such that: $c_{0.95} \simeq 1.358$.

Remark 5.3.1. This is remarkable that we can use a framework in which all parameters are known, even though we actually do not know the parameter λ beforehand.

Note that we may have ties here in practice, because we kept a day precision in the times of arrival. Therefore, we cannot compute D_n exactly. Neither can we use the usual critical values. Noether (1963) proved that the test is conservative in the discrete case: if we reject the null hypothesis with our estimated D_n for the standard critical values corresponding to a $1-\alpha$ confidence, then the actual confidence would be greater or equal. Methods exist to use Kolmogorov-Smirnov (KS) tests with grouped data (Pettitt and Stephens, 1977). Yet, our objective is to show that our assumption is reasonable, and not to provide accurate discriminatory tests.

Remark 5.3.2. One may argue that we should have kept a minute precision in the arrival of donors to avoid such problems. We think on the opposite that the minute precision was not relevant. We use the time of death of the donor as an arrival time for the donor and assume that all kidney offers are punctual in time (assumption 4.1.1). This assumption is reasonable

at a day precision. Figure 5.5 gives evidence that the day granularity is reasonable because only 0.5 donor arrives each day in average. Furthermore, we are not interested in the fact that the death of donors is Poisson inside the same day.

Test

We use the test on our data. We give in figure 5.6 the KS curves, comparing the empirical and the theoretical CDFs. Again, this confirms that the donor arrival rate was smaller in 2012 than in 2017, as the empirical CDF is globally under the expected CDF.

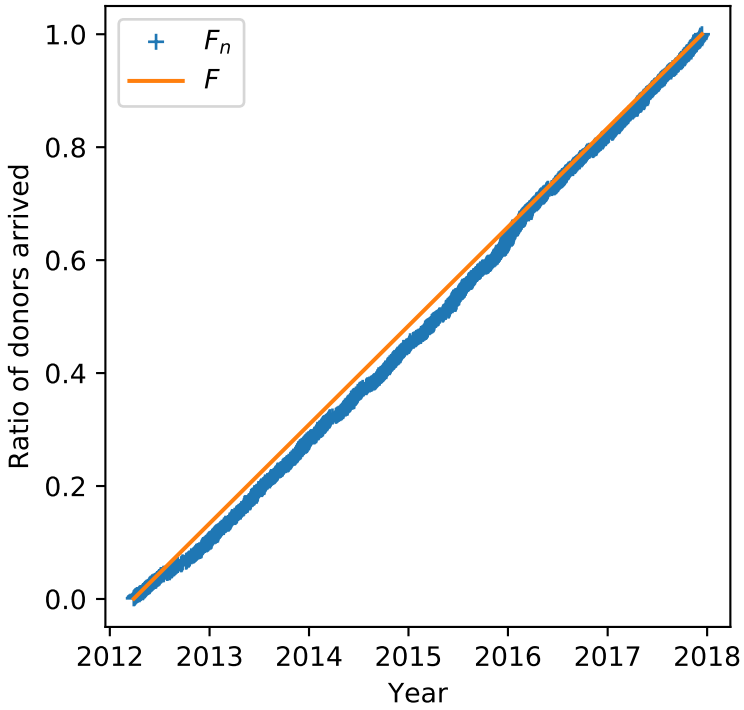

Figure 5.6 Empirical vs. theoretical CDF of donors arrival between 2012-03-29 and 2017-12-13.

We computed the KS adjusted statistic in 5.12 for different time windows. The null hypothesis is that the arrival of donors is a Poisson point process. For each of the two-year time windows, we cannot reject the null hypothesis even with an as low as 85% confidence ($c_{0.85} \simeq 1.138$), assuming our statistic is a good estimate of the actual one. However, we could reject the null hypothesis for the whole study with 90% confidence ($c_{0.9} \simeq 1.224$). This

brings support to our intuition that the whole process is not a constant Poisson but varies softly with time.

However, one should be careful while interpreting the KS test, as it is meant to reject the null hypothesis, and not to accept it. Incidentally, we do not wish to statistically "prove" that our process is Poisson as we know that it never is in reality. We only want to support the assumption that it is a good model.

5.3.2 Probability Plots

To provide additional evidence for our assumption, we propose to use the well-known Quantile-Quantile (QQ)-plot, as a visual way of convincing the reader. We rely on the methodology explained by Law (2015) p. 339-343. Here, there is no problem with ties, as QQ-plots allow for tie management. As we are working with a theoretical uniform distribution, the QQ-plot displays the same information as the PP-plot (with only a change in axis labelling: which scales from 0 to 1 for the PP-plot). We will prefer showing the PP-plot because we feel it is more straightforward to understand.

We give the PP-plot for the whole study in figure 5.7. The interpretation of the plot is qualitative, but brings support to assumption 4.1.5 in our opinion. The small gap to the first bisector for intermediate probabilities is due to the increase in donors arrival rate. This vanishes with a smaller time window.

Remark 5.3.3. In spite of the apparent similarity between figures 5.6 and 5.7, they are inherently different. The presence of the first bisector in both figures is due to the fact that our null hypothesis is a uniform distribution.

Table 5.12 Kolmogorov-Smirnov adjusted statistic estimate for different time windows.

Time Window	KS Adjusted Statistic Estimate
2012-12-13 to 2014-12-13	0.439
2013-12-13 to 2015-12-13	0.892
2014-12-13 to 2016-12-13	0.765
2015-12-13 to 2017-12-13	0.887
2012-03-29 to 2017-12-13	1.308

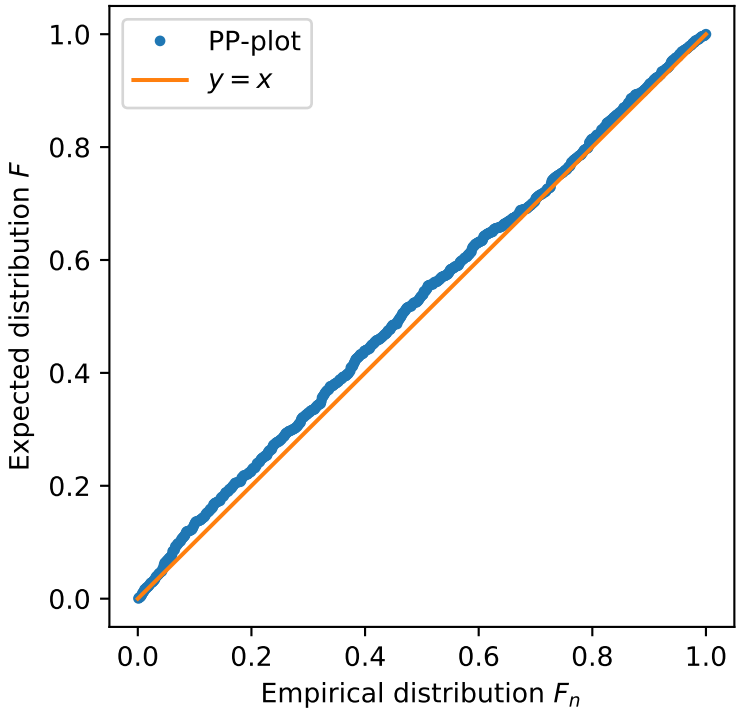

Figure 5.7 PP plot of donors arrival between 2012-03-29 and 2017-12-13 vs. first bisector.

5.4 Training, Validation and Test Sets

The following section presents how we split the data of TQ into validation and test set to validate the algorithm presented in section 4.3.

5.4.1 Training Set

We want to clarify what we call training set here. In usual supervised machine learning problems, the training set is a set of observations with a known target. For this set of observations, a loss function is minimised which measures the distance between predictions and targets. In our problem, we do not have such a training set. Our training set is composed of donors arrived in the past ΔT days (see section 4.3.1). To make the prediction of the next kidney offer, we do not use the target or any information available in the future. For each patient-donor offer, the training set is different, because it makes use of the latest arrived donors.

5.4.2 Validation and Test Sets

Building Targets

To verify our algorithm, we have to compare our predictions to the observations. We do the following operations to build the datasets:

- We select all recovered donors who go up to the general attribution list.
- We remove all offers for candidates with a special priority.
- We remove all "artificial" offers.
  Each time a transplantation centre refuses a kidney for all of its patients, they are all considered as having had an offer. That is not what we want, as patients did not have to make a choice for these "offers".
- We remove all incompatible offers.
  When a patient is found to be incompatible after additional cross-match tests, we omit the offer.
- For non-dialysed patients, we arbitrarily selected the date of the offer as a waiting-date. The rationale under it is the following: if the patient has a high priority on the scoring-list, he will have an offer in a short time; if the patient has a low priority on the scoring-list, he will probably start dialysis before getting an offer. In the first case, considering a potential increase in waiting-score will have little consequence on the predicted distribution. In the second case, considering a potential increase in waiting-score will avoid over-estimating the time of next offer.
- For each offer to a patient we retrieve the target.
  If the next event was an offer, we save the time to the next offer.
  If the next event was not an offer, we save the time to the next event: permanent or temporary removal from waiting-list (including living donor transplantation). We mark the observation as censored.
- We remove candidates with incoherent waiting-times or times of first dialysis (negative time between the aforementioned time and the first offer).

Finally, we have 2365 observations corresponding to 1757 patients.

There is a final step to do before splitting the dataset. For each observation, we need a training set, i.e. a pool of donors ΔT days backwards in time. So far, the arrival dates scale from 2012-03-29 to 2017-12-13. We have to remove the donors in the ΔT first days of the study.

In practice, we took $\Delta T = 730$, according to the considerations in 5.3. Even though we might use a smaller time (e.g. $\Delta T = 365$) for some experiments, we will not change our sets, for comparison purposes. Of course, it would be valuable to keep the same time space forward in time for every patients, to test our algorithm on future pool of donors but this would reduce the size of the sets too much.

The inter-offers times are distributed as showed in 5.8. The mean time between two consecutive offers for the same patient (as defined above) is 78.5 days, the median is 28 days and the standard deviation 136.7 days. This means half of offers lead to another offer within one month.

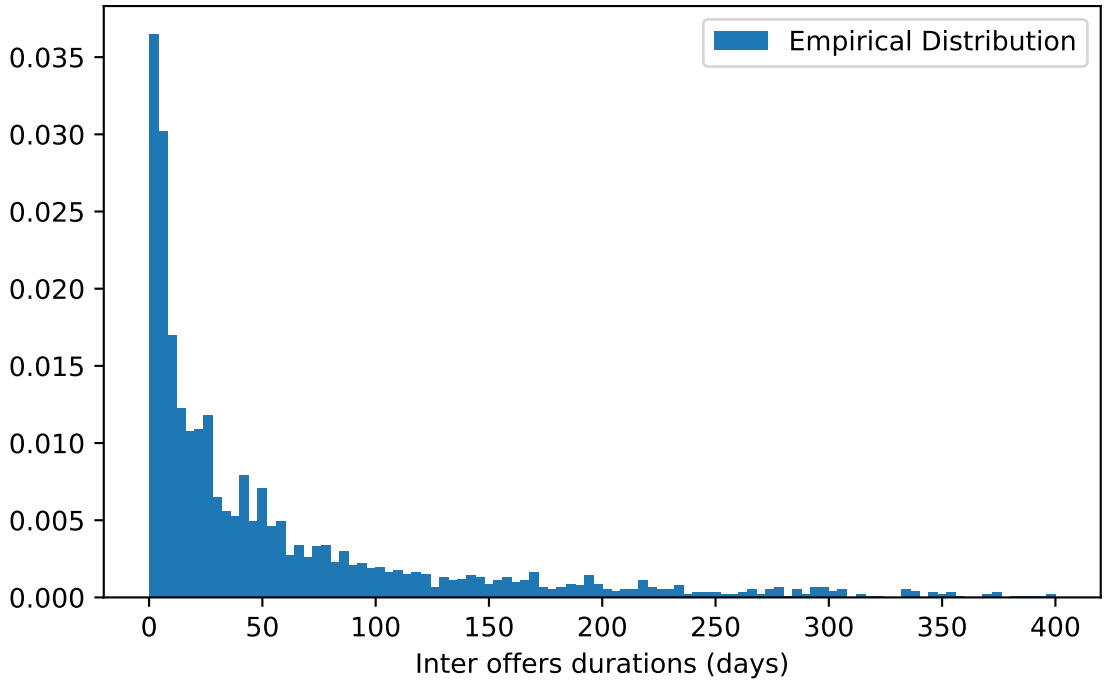

Figure 5.8 Distribution of the inter-offers-times across the kidney offers made in Québec between 2012-03-29 and 2017-12-13.

From this set, we removed the candidates for which there was a risk of wrong calculation of the score, not to bias our predictions. From the list of candidate-donor scores mismatches (errors in the score that we computed in comparison to the one of TQ), we decided to remove all candidates with a difference in the score greater than 0.05 (only keeping small errors due to age).

Splitting the Dataset

Now, we split the dataset between a validation and a test set. We did a 50/50 random splitting, respecting the proportion of censored and uncensored values. In the end, we have two sets each with 712 observations as explained in table 5.13.

Table 5.13 Validation and test sets. The same candidate can have multiple offers.

Set	Uncensored	Censored	Total	Number of Candidates
Validation	569	143	712	414
Test	569	143	712	417

Remark 5.4.1. There is a special relation between training set and validation set. As we mentioned, there is a difference in the nature of the sets. However, strictly speaking, those sets are not independent. In the training set, we use information about the decisions of patients before the current offer, which had an influence on which patients got an offer. This means that we indirectly use previous offers to make predictions for the current offer, even though these previous offers are in the validation set. Moreover, the offers in the validation set are not independent of each other, strictly speaking. Indeed, the same patient might have had several offers, so we can find several offers for the same patient in the validation set. We neglect all these subtleties in the rest of the work for convenience. The quality of our results will implicitly confirm our hypotheses.

Remark 5.4.2. The careful reader may have stumbled on the title of this section: "Observational Data". Of course, it is observational because it is not simulated. Beyond this, the fact that our data is observational has several consequences on our validation and test datasets. Our verifications will be more significant for categories of patients likely to get more offers (because they will have been averaged over more occurrences) than for patients who rarely have offers. It is very logical to have less examples for low-priority or hard-to-match patients but as those patients present the greatest inherent stochasticity, it will be difficult to assess whether the algorithm performs correctly on these. The inclusion of censored observations partly counterbalances this phenomenon, because censored observations are more likely to originate from low-priority patients.

CHAPTER 6 VERIFICATIONS

We verified the algorithms predicting the time to next offer on the datasets which we introduced in section 5.

6.1 Verification Methodology

Validation is difficult in our problem. There is not one single cost function to simultaneously rank our models and quantify the quality of the predictions. Instead, we will have to qualitatively collect the information from different cost functions and choose subjectively the best model. For each model, from a clinical perspective, we would like to be able to evaluate the absolute quality of the predictions, to define "bad" predictions and measure their impact.

6.1.1 Foreword

The distribution of incoming eligible donors is unique for each patient at each time. In order to validate the parameters that we estimate and the times that we deduce from it, we only have one observation at our disposal: the actual time to next offer (or a time of censoring). The temporal predictions that we can make from the distribution are the expected value $\mathbb{E}(T)$ and the quantiles t_α which we cannot *a priori* compare to the actual value t. In a usual problem, we would have several observations from one distribution and average them to get an estimate of the real observed expected value to compare to the predicted expected value.

One would think of the usual statistical tools to measure the quality of the predictions: MSE or Mean Absolute Percentage Error (MAPE). Why is our situation different from the typical machine learning situation? In the common machine learning or statistical approaches, one tries to minimise a loss function over a training set with certain targets (the actual times to next offer). Yet, here, we are not using the actual times to next offer in the "training set", and additionally the "training set" that we use here is of a different nature (current patient, current donor, past donors, past waiting-list and current waiting-list) than our "validation and test sets", which are made of patient/donor/time to next offer triplets. In our algorithm, we inherently did not aim at learning the times to next offer, but the distribution of times to next offer. Consequently, we will see how to make good use of the traditional measures to compare the different variants of our algorithm and we will develop other measures to capture the absolute predictive quality of these variants.

We will distinguish three types of measures:

- "Relative" measures: these measures aim at giving indexes to choose between different algorithms/sets of hyper-parameters. We do not use the word "discriminatory" here not to cause confusions with the survival terminology.
- "Absolute" measures: these measures aim at capturing the quality of prediction and the absolute performances of the model. They are not meant to compare models in the first place. This is crucial in medicine: we do not want to select the "less worse" algorithm/set of hyper-parameters among the ones we developed, we want to see if the algorithm makes good predictions which can be reasonably conveyed to the patients.
- Detectors of "bad predictions": these measures aim at identifying which predictions are "bad", in a sense which we will define, in order to understand their cause and study the seriousness of these predictions. This is also crucial in medicine, because we cannot afford to mislead a patient at the individual level.

Remark 6.1.1. In the literature, most papers do not address these concerns because they stand at a more political level, at which the main preoccupation is the overall social welfare.

Additionally, it is important we should think about taking censored values into account, as we may otherwise bias our estimation. Censoring is an important notion in survival analysis. In our problem, a patient is censored if he died or was removed from the waiting-list before the next offer. The only information on the time to next offer is that it is greater than the time of censoring. There is undoubtedly a correlation between the time to next offer and the probability of an event: the more a patient has to wait for a transplant, the more likely he is to become too sick to transplant or to die (even though the scoring function used to rank patients on waiting-list is meant to attenuate those effects). Therefore we try to assess the possibility of including censored values in our validating methods.

Notations

To avoid confusions with section 4, we call $T_{(i)}$ the time to next offer corresponding to a certain offer (x_i, y_i), whereas T_i is the ith donor arrived on the general attribution list.

Additionally, we add the exponent c to refer to a time of censoring: T^c instead of T for instance.

6.1.2 Mean Squared Error and Mean Absolute Percentage Error

Mean Squared Error

The MSE is a well-known measure of quality of a predictor.

Definition 6.1.1 (Mean Squared Error)**.** Let X_i be independent random variables of predicted value $\hat{X}_i$. The Mean Squared Error is:

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \hat{X}_i\right)^2$$

In our problem, it writes:

$$\frac{1}{n}\sum_{i=1}^{n}\left(T^*_{(i)} - \mathbb{E}(T^{\simeq}_{(i)})\right)^2$$

This is an interesting relative measure. However, it penalises mistakes in long-term prediction more than short-term: an error of one week for a high-priority patient will be less penalised than an error of one month for a low-priority patient, even though a one month precision can be very good for a low-priority patient and a one week precision poor for a high-priority patient, due to the high stochasticity of low-priority patient's offers. In what follows, one of our concern will be to find an error which disentangles inherent stochasticity of the prediction and quality of the prediction. We do not want to penalise stochasticity but the error of prediction.

Mean Absolute Percentage Error

The MAPE is also a well-known measure of quality of a predictor.

Definition 6.1.2 (Mean Absolute Percentage Error)**.** Let X_i be independent random variables of predicted value $\hat{X}_i$. The Mean Absolute Percentage Error is:

$$\frac{100\%}{n}\sum_{i=1}^{n}\frac{\left|X_i - \hat{X}_i\right|}{X_i}$$

In our problem, it writes:

$$\frac{100\%}{n}\sum_{i=1}^{n}\frac{\left|T^*_{(i)} - \mathbb{E}(T^{\simeq}_{(i)})\right|}{T^*_{(i)}}$$

One advantage for our problem is that it would not over-penalise errors for large observed times.

This error is not fully satisfying though as it may give too big penalties for little observations which were legitimately predicted bigger. Moreover, there is a problem when the observed next offer took place the same day: which results in a zero denominator. Of course, it is always possible to use $\max(1, T_{(i)}^{observed})$ as a denominator. This may be corrected using other measures such as the Symmetric Mean Absolute Percentage Error (SMAPE) (Flores, 1986):

$$\frac{100\%}{n}\sum_{i=1}^{n} 2\frac{\left|X_i - \hat{X}_i\right|}{|X_i| + |\hat{X}_i|}$$

6.1.3 Mean Normalised Squared Error

We propose a new error, taking advantage of the fact that we predict the distribution of the time to next offer. Instead of normalising by the observed time as for the MAPE, we will normalise by the predicted variance. We will see this might be a misleading relative measure, but that it can be used successfully as a means of identifying bad predictions. We introduce at first the Mean Normalised Squared Indicator:

Definition 6.1.3 (Mean Normalised Squared Indicator)**.** Let X_i be independent random variables of expected value μ_i and variance σ_i^2. The Mean Normalised Squared Indicator is:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \mu_i}{\sigma_i}\right)^2$$

Remark 6.1.2. The rationale behind this indicator is that: $\mathbb{E}\left(\frac{X_i - \mu_i}{\sigma_i}\right) = 0$ and $\mathrm{Var}\left(\frac{X_i - \mu_i}{\sigma_i}\right) = 1$. It makes the different observations "comparable" by normalising them.

This indicator is not an error as such. We turn it into a Mean Normalised Squared Error (MNSE) by replacing the actual expected values and variances by the predicted ones. In our problem, the MNSE is: $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left(T_{(i)} - \mathbb{E}(T_{(i)})\right)^2}{\mathrm{Var}(T_{(i)})}\right)$. In practice, it will write as:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left(T_{(i)}^{*} - \mathbb{E}(T_{\widetilde{(i)}})\right)^2}{\mathrm{Var}(T_{\widetilde{(i)}})}\right)$$

A prediction will be particularly penalised if it is little (thus little variance) whereas the observation is actually big. This is what we expect, as we do not want to make a patient wait in vain for a kidney transplant, after declining the current offer. Moreover, it will penalise also big predictions for which the observation is little and very unlikely in our model, as the variance will not be big enough to compensate the deviation. Still, this indicator might overestimate the performance of a model being very conservative but also very inaccurate (always giving too big predictions).

In other words, this error is reliable under the null hypothesis, i.e. assuming that our prediction is already good. This is problematic if we want to use this error to compare different sets of hyper-parameters. However, this can be turned into a relevant measure to compare sets of hyper-parameters which have already been validated through other methods.

Under the null hypothesis, the NSE is a way of defining bad predictions:

Definition 6.1.4 (Bad Prediction)**.** A bad prediction is a prediction which results in a NSE above a given threshold.

The threshold for bad predictions depends on the method and on the data.

Remark 6.1.3. One may argue against our terminology of "bad predictions", because the word "bad" is relative and subjective. However, we prefer this term in comparison to others like inaccurate, wrong or incorrect. It is not possible to know with certainty when a prediction is wrong, because the observed time to next offer is subject to stochasticity. Unlikely would be a better term, but it is already dedicated to the likelihood of the prediction. Thus the word "bad" encapsulates the fact that this notion is definition-dependent and also intuitive.

Remark 6.1.4. We chose not to use the log likelihood in our problem for two main reasons. The log-likelihood is too sensitive to little observed times which were predicted as big. Indeed, due to the way our algorithm works, it is common that the predicted probability of getting an offer in the first months after the current offer is zero, because no donor was found eligible in the training set. This does not mean that the prediction is necessarily bad: such situations occur when an exceptional match for the patient arrives, which could not have been predicted looking at the previous donors. Even if the likelihood was zero, the expected time itself is not infinite, as it accounts for future evolutions of the patient's score. Maybe the difference between predicted expected time and observed time is not that dramatic. The second reason is that the different values of the likelihood were not comparable (leading to the same problem as MSE): the predicted distributions are continuous, but the density function has not the

same shape or maximum value according to the priority of patients. We do not want an indicator to confuse inherent stochasticity with bad predictions.

6.1.4 Concordance Index

A well-known indicator to evaluate the relative predictive power of a model in survival analysis is the Concordance Index (or C-Index, or C-Statistic, often named Harrell's C-Index in tribute to its inventor). It measures the ability of a survival model to predict the events in the right order: if an event happened to a patient before another patient, then the predictions should be in the same order. This is an interesting indicator as it accounts for censored as well as uncensored data.

Definition 6.1.5 (C-index in the continuous case)**.** Let t_i be observed times for different patients i, $\hat{t}_i$ the predicted times and $C_i = 0$ if the observation is censored and 1 otherwise. We assume t_i and $\hat{t}_i$ are continuous. The C-index is:

$$C^{index} = \frac{\sum_{i,C_i=1} \sum_{j \neq i} \mathbb{1}_{t_i<t_j} \mathbb{1}_{\hat{t}_i<\hat{t}_j}}{\sum_{i,C_i=1} \sum_{j \neq i} \mathbb{1}_{t_i<t_j}}$$

Remark 6.1.5. If the times are not assumed to be continuous, there exists some special way of handling ties (i.e. equality). We will use this version in practice, as implemented in the Python package lifelines 0.12.0.

In our problem, we use $\mathbb{E}(T^{\approx}_{(i)})$ as the predicted time for the C-index.

We will use the C-index both as an absolute and a relative measure of the performance for the algorithm. We feel it is appropriate in our context. The performance that it captures is the significance of differences in predicted expected values. However, it does not capture the absolute quality of the predicted time as we will see: a model which systematically overestimates the waiting-times could have a very good C-index. We can see our problem as a survival problem: the time to next offer is the time of death and death/removal are times of censoring. Pencina and D'Agostino (2004) discuss the difference between discrimination and calibration of a survival model: the former is the ability to rank patients and the latter the fit between predicted probabilities and observed probabilities. C-index is in this survival terminology a discriminatory measure. For a deeper interpretation of the C-index and its relation to the AUROC we refer to Pencina and D'Agostino (2004); Austin and Steyerberg (2012).

6.1.5 Wasserstein Distance

As stated before, the verification is difficult here, because we want to verify a whole distribution with a single observation. This is exactly the purpose of the Wasserstein distance, also called "earth mover" distance (Wasserstein, 1969). It is meant to compare two different distributions seen as two piles of earth which you want to merge into one pile by minimising the cost of bringing each quantum of density to another place. In our problem, we want to compare the predicted distribution to a dirac, representing the observed time.

Definition 6.1.6. (Wasserstein distance to a dirac)
Let d be a distance in $\mathbb{R}$. Let $x_0 \in \mathbb{R}_+$ of probability measure δ_0 and X an integrable random variable of probability measure f. The p^{th} Wasserstein distance between δ_0 and f is:

$$W_p(f, \delta_0) = \left(\int_{\mathbb{R}} d(x, x_0)^p f(x) dx \right)^{\frac{1}{p}} = \mathbb{E}_X \left(d(X, x_0)^p \right)^{\frac{1}{p}}$$

Let us take $d = |\cdot|$. Let $T^*_{(i)}$ by the actual observed time to next offer for offer i associated to dirac δ_i (considered as deterministic here). We call g_i the predicted density function of $T^{\simeq}_{(i)}$.

Lemma 6.1.1.

(i) The first Wasserstein distance in our problem is:

$$W_1(g_i, \delta_i) = T^*_{(i)}(1 - 2\mathbb{P}(T^{\simeq}_{(i)} \geq T^*_{(i)})) - \mathbb{E}(T^{\simeq}_{(i)}) + 2\mathbb{E}(T^{\simeq}_{(i)} | T^{\simeq}_{(i)} \geq T^*_{(i)})\mathbb{P}(T^{\simeq}_{(i)} \geq T^*_{(i)}),$$

(ii) The second Wasserstein distance in our problem is:

$$W_2(g_i, \delta_i) = \left(\text{Var}(T^{\simeq}_{(i)}) + (\mathbb{E}(T^{\simeq}_{(i)}) - T^*_{(i)})^2 \right)^{\frac{1}{2}}$$

We have closed formulas for both distances.

Remark 6.1.6. The above probabilities, expected values and variances refer to the predicted distribution.

Proof. The calculation is straightforward and relies on:

$$W(g_i, \delta_i) = \int_0^{+\infty} |t - T^*_{(i)}| g_i(t) dt = \int_0^{T^*_{(i)}} (T^*_{(i)} - t) g_i(t) dt + \int_{T^*_{(i)}}^{+\infty} (t - T^*_{(i)}) g_i(t) dt ...$$

□

This distance can be used as a relative measure between models. However, looking at the formulas we obtain, we think that it has more the structure of a loss function for the training of a ML algorithm. For example, the second Wasserstein distance penalises the stochasticity, which is useful to minimise a loss but not necessarily relevant when comparing models. This is the result of assuming that the observed distribution is a dirac. The first distance "tries to learn" the median, whereas the second distance "tries to learn" the expected value, for these distributions.

From the Wasserstein distance, we define the Mean Wasserstein Distance (MWD), as the average of the Wasserstein distances over the dataset.

6.1.6 Empirical Mean Quantiles

The section below is a methodological contribution of this work, to the best of our knowledge. The idea is to estimate the proportion of patients p_α in our dataset such that $t \leq t_\alpha$ for each desired α. If the model is correct, we should get $p_\alpha \simeq \alpha$. This is not as obvious as it seems, because we consider independent random variables but do not assume they are identically distributed.

Theorem 6.1.1 (Empirical Mean Quantiles)**.** *Let $n \in \mathbb{N}^*$.*
Let $(T_{(i)})$ be a series of independent random variables (typically describing the time to next offer for a (x_i, y_i, z_i) triplet (patient, donor, observed time to next offer)). $\forall i \in \{1, ..., n\}$, $T_{(i)} \sim \mathcal{T}_i$.
Let $t_{i,\alpha}$ be the quantiles such that $\mathbb{P}(T_{(i)} \leq t_{i,\alpha}) = \alpha$ and $B_i = \mathbb{1}_{T_{(i)} \leq t_{i,\alpha}}$, $\forall i$. Then:

$$\bar{B}_n = \frac{\sum_{i=1}^{n} B_i}{n} \overset{a.s.}{\to} \alpha$$

And with 95% probability, provided n is big enough:

$$\alpha \in \left[\bar{B}_n - 1.96\sqrt{\frac{\bar{B}_n(1-\bar{B}_n)}{n}}, \bar{B}_n + 1.96\sqrt{\frac{\bar{B}_n(1-\bar{B}_n)}{n}} \right]$$

Proof. Let $\alpha \in [0, 1[$.

Let $\mathcal{T}_i$ be the distribution of $T_{(i)}$, random variable describing the time to next offer, for a (x_i, y_i, z_i) triplet (patient, donor, observed time to next offer). We assume that the $T_{(i)}$ are independent from each other. As the $\mathcal{T}_i$ may differ from each other, the $T_{(i)}$ are not identically distributed *a priori*.

We define the quantiles $t_{i,\alpha}$ such that $\mathbb{P}(T_{(i)} \leq t_{i,\alpha}) = \alpha$.

Let $B_i = \mathbb{1}_{T_{(i)} \leq t_{i,\alpha}}$, $\forall i$. We have that $\mathbb{P}(B_i = 1) = \mathbb{P}(T_{(i)} \leq t_{i,\alpha}) = \alpha$. So $B_i \sim \mathcal{B}(\alpha)$, $\forall i$. And thus the B_i's are independent and identically distributed. We can apply the Law of Large Numbers so that:

$$\frac{\sum_{i=1}^{n} B_i}{n} \overset{a.s.}{\to} \alpha$$

The confidence intervals follow immediately from the Central Limit Theorem. □

In practice, we will estimate:

$$\bar{B}_n = \frac{\sum_{i=1}^{n} \mathbb{1}_{T^*_{(i)} \leq \tilde{t}_{i,\alpha}}}{n}$$

This could be turned into a statistical test. We would reject the null hypothesis when the theoretical α is not in the estimated confidence intervals for a certain confidence. However, we prefer to use it as a qualitative way of representing the quality of the fitting of the model to the data. Typically, it is important that the extreme quantiles should be accurate (e.g. $\tilde{t}_{i,10\%}$ and $\tilde{t}_{i,95\%}$) because this validates the model's calibration: neither over-estimating nor under-estimating.

Remark 6.1.7. Unfortunately, we were not able to include censored values in this framework. All our attempts required assumptions about the distribution of time of censoring.

6.1.7 Local means

To the best of our knowledge, what we present below is a methodological contribution about handling censored values from multiple estimated distributions. We use the predicted distributions and times of censoring to infer the unobserved times to next offer. Then we cluster the observations by same predicted expected time to next offer. If the model is correct, we show that the actual observed times should be equal on average to the predicted time.

Unobserved Times Estimation

When an observation is censored, the only information that we have on the real time to next offer is the time of censoring. Therefore, we would like to draw all possible information from this censoring time to estimate the actual time to next offer. Let us assume that we

know exactly the distribution of T (time to next offer). The most logical estimate for T knowing T^c is derived from the the so-called mean residual life: $\mathbb{E}_T(T|T > T^c)$.

Now, we would like to use $\mathbb{E}_T(T|T > T^c)$ as an observation when we do not know T. We achieve this in practice by making the following assumption:

Assumption 6.1.1. *$T^{\simeq} \sim T^*$ (null hypothesis). So $\mathbb{E}(T^{\simeq}|T^{\simeq} > T^c, T^c) = \mathbb{E}(T^*|T^* > T^c, T^c)$.*

Remark 6.1.8. In other words, $\mathbb{E}_T(T|T > T^c)$ can be seen as the estimation of the counterfactuals. The counterfactual being "what would have happened if what happened had not happened".

Let $\bar{T} = \mathbb{1}_{T \leq T^c} T + \mathbb{1}_{T > T^c} \mathbb{E}_T(T|T > T^c)$. $\bar{T}$ is the random variable standing for the randomness of censoring of an observation. We recall that T and T^c are independent. We have the following theorem:

Theorem 6.1.2 (Randomly censored value estimate)**.**

(i) $\mathbb{E}(\bar{T}) = \mathbb{E}(T)$

(ii) $\mathrm{Var}(\bar{T}) = \mathrm{Var}(T) - \mathbb{P}_{T,T^c}(T > T^c)\,\mathrm{Var}_{T,T^c}(T|T > T^c)$

(iii) $\forall k \geq 1$, *as long as the moments exist:* $\mathbb{E}\left(|\bar{T} - \mathbb{E}(\bar{T})|^k\right) \leq \mathbb{E}\left(|T - \mathbb{E}(T)|^k\right)$

Proof. The three proofs follow the same scheme: calculation of the moment conditioned by T^c, integration over T^c and law of total expectation.

(i)

$$
\begin{aligned}
\mathbb{E}(\bar{T}|T^c) &= \mathbb{P}(T \leq T^c|T^c)\mathbb{E}(\bar{T}|T \leq T^c, T^c) + \mathbb{P}(T > T^c|T^c)\mathbb{E}(\bar{T}|T > T^c, T^c), \\
&\qquad \text{law of total expectation} \\
&= \mathbb{P}(T \leq T^c|T^c)\mathbb{E}(T|T \leq T^c, T^c) + \mathbb{P}(T > T^c|T^c)\mathbb{E}(\mathbb{E}(T|T > T^c, T^c)|T > T^c, T^c) \\
&= \mathbb{P}(T \leq T^c|T^c)\mathbb{E}(T|T \leq T^c, T^c) + \mathbb{P}(T > T^c|T^c)\mathbb{E}(T|T > T^c, T^c) \\
&= \mathbb{E}(T|T^c), \text{ law of total expectation} \\
&= \mathbb{E}(T), \text{ because } T \text{ and } T^c \text{ independent}
\end{aligned}
$$

Finally, $\mathbb{E}(\bar{T}) = \mathbb{E}(\mathbb{E}(\bar{T}|T^c)) = \mathbb{E}(\mathbb{E}(T)) = \mathbb{E}(T)$

(ii)

$$\begin{aligned}
\mathrm{Var}(\bar{T}|T^c) &= \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left((\bar{T}-\mathbb{E}(T))^2\Big|\, T \le T^c, T^c\right) \\
&\quad + \mathbb{P}(T > T^c|T^c)\mathbb{E}\left((\bar{T}-\mathbb{E}(T))^2\Big|\, T > T^c, T^c\right), \\
&\quad \text{law of total expectation and } \mathbb{E}(\bar{T}) = \mathbb{E}(T) \\
&= \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left((T-\mathbb{E}(T))^2\Big|\, T \le T^c, T^c\right) \\
&\quad + \mathbb{P}(T > T^c|T^c)\mathbb{E}\left((\mathbb{E}(T|T > T^c, T^c)-\mathbb{E}(T))^2\Big|\, T > T^c, T^c\right) \\
&= \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left((T-\mathbb{E}(T))^2\Big|\, T \le T^c, T^c\right) \\
&\quad + \mathbb{P}(T > T^c|T^c)(\mathbb{E}(T|T > T^c, T^c)-\mathbb{E}(T))^2 \\
&= \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left((T-\mathbb{E}(T))^2\Big|\, T \le T^c, T^c\right) + \mathbb{P}(T > T^c|T^c) \\
&\quad \times \left(\mathbb{E}(T|T > T^c, T^c)^2 + \mathbb{E}(T)^2 - 2\mathbb{E}(T)\mathbb{E}(T|T > T^c, T^c)\right)
\end{aligned}$$

We make $\mathbb{E}(T^2|T > T^c, T^c)$ appear in the expression:

$$\begin{aligned}
\mathrm{Var}(\bar{T}|T^c) &= \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left((T-\mathbb{E}(T))^2\Big|\, T \le T^c, T^c\right) + \mathbb{P}(T > T^c|T^c) \\
&\quad \times \left(\mathbb{E}\left(T^2\Big|\, T > T^c, T^c\right) - \mathbb{E}\left(T^2\Big|\, T > T^c, T^c\right) + \mathbb{E}\left(T|T > T^c, T^c\right)^2\right. \\
&\quad \left. +\mathbb{E}(T)^2 - 2\mathbb{E}(T)\mathbb{E}(T|T > T^c, T^c)\right) \\
&= \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left((T-\mathbb{E}(T))^2\Big|\, T \le T^c, T^c\right) + \mathbb{P}(T > T^c|T^c) \\
&\quad \times \left(-\mathrm{Var}(T|T > T^c, T^c) + \mathbb{E}\left(T^2\Big|\, T > T^c, T^c\right) + \mathbb{E}\left(\mathbb{E}(T)^2\Big|\, T > T^c, T^c\right)\right. \\
&\quad \left. -2\mathbb{E}\left(T\mathbb{E}(T)|T > T^c, T^c\right)\right) \\
&= \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left((T-\mathbb{E}(T))^2\Big|\, T \le T^c, T^c\right) + \mathbb{P}(T > T^c|T^c) \\
&\quad \times \left(-\mathrm{Var}(T|T > T^c, T^c) + \mathbb{E}\left(T^2 + \mathbb{E}(T)^2 - 2T\mathbb{E}(T)\Big|\, T > T^c, T^c\right)\right) \\
&= \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left((T-\mathbb{E}(T))^2\Big|\, T \le T^c, T^c\right) + \mathbb{P}(T > T^c|T^c) \\
&\quad \times \left(-\mathrm{Var}(T|T > T^c, T^c) + \mathbb{E}\left((T-\mathbb{E}(T))^2\Big|\, T > T^c, T^c\right)\right) \\
&= \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left((T-\mathbb{E}(T))^2\Big|\, T \le T^c, T^c\right) \\
&\quad + \mathbb{E}\left((T-\mathbb{E}(T))^2\Big|\, T > T^c, T^c\right)\mathbb{P}(T > T^c|T^c) \\
&\quad - \mathbb{P}(T > T^c|T^c)\,\mathrm{Var}(T|T > T^c, T^c) \\
&= \mathbb{E}\left((T-\mathbb{E}(T))^2\Big|\, T^c\right) - \mathbb{P}(T > T^c|T^c)\,\mathrm{Var}(T|T > T^c, T^c), \\
&\quad \text{law of total expectation} \\
&= \mathrm{Var}(T|T^c) - \mathbb{E}\left((T-\mathbb{E}(T|T > T^c))^2\mathbb{1}_{T>T^c}\Big|\, T^c\right)
\end{aligned}$$

Finally, by integrating over T^c:

$$\begin{aligned}\mathrm{Var}(\bar{T}) &= \mathbb{E}(\mathrm{Var}(\bar{T}|T^c)) \\ &= \mathrm{Var}(T) - \mathbb{E}_{T,T^c}\left((T - \mathbb{E}(T|T > T^c))^2 \mathbb{1}_{T>T^c}\right) \\ &= \mathrm{Var}(T) - \mathbb{P}_{T,T^c}(T > T^c)\,\mathrm{Var}_{T,T^c}(T|T > T^c)\end{aligned}$$

(iii)

$$\begin{aligned}\mathbb{E}\left(|\bar{T} - \mathbb{E}(\bar{T})|^k \middle| T^c\right) &= \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left(|\bar{T} - \mathbb{E}(T)|^k \middle| T \le T^c, T^c\right) \\ &\quad + \mathbb{P}(T > T^c|T^c)\mathbb{E}\left(|\bar{T} - \mathbb{E}(T)|^k \middle| T > T^c, T^c\right), \\ &\text{law of total expectation and } \mathbb{E}(\bar{T}) = \mathbb{E}(T) \\ &= \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left(|T - \mathbb{E}(T)|^k \middle| T \le T^c, T^c\right) \\ &\quad + \mathbb{P}(T > T^c|T^c)\mathbb{E}\left(|\mathbb{E}(T|T > T^c, T^c) - \mathbb{E}(T)|^k \middle| T > T^c, T^c\right) \\ &= \mathbb{P}(T \le T^c|T^c)\mathbb{E}(|T - \mathbb{E}(T)|^k \middle| T \le T^c, T^c) \\ &\quad + \mathbb{P}(T > T^c|T^c)|\mathbb{E}(T|T > T^c, T^c) - \mathbb{E}(T)|^k \\ &= \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left(|T - \mathbb{E}(T)|^k \middle| T \le T^c, T^c\right) \\ &\quad + \mathbb{P}(T > T^c|T^c)|\mathbb{E}(T - \mathbb{E}(T)|T > T^c, T^c)|^k \\ &\le \mathbb{P}(T \le T^c|T^c)\mathbb{E}\left(|T - \mathbb{E}(T)|^k \middle| T \le T^c, T^c\right) \\ &\quad + \mathbb{P}(T > T^c|T^c)\mathbb{E}\left(|T - \mathbb{E}(T)|^k \middle| T > T^c, T^c\right), \\ &\text{by Jensen's inequality with } x \mapsto |x|^k \text{ convex with } k \ge 1 \\ &= \mathbb{E}\left(|T - \mathbb{E}(T)|^k \middle| T^c\right), \text{ law of total expectation} \\ &= \mathbb{E}\left(|T - \mathbb{E}(T)|^k\right)\end{aligned}$$

Finally, by monotonicity of the expected value:

$$\begin{aligned}\mathbb{E}\left(|\bar{T} - \mathbb{E}(\bar{T})|^k\right) &= \mathbb{E}\left(\mathbb{E}\left(|\bar{T} - \mathbb{E}(\bar{T})|^k|T^c\right)\right) \\ &\le \mathbb{E}\left(\mathbb{E}\left(|T - \mathbb{E}(T)|^k\right)\right) \\ &= \mathbb{E}\left(|T - \mathbb{E}(T)|^k\right)\end{aligned}$$

□

Local Means with Censored Values

We cluster the observations according to their estimated expected value, to obtain groups of similarly distributed observations. In practice, as these expected values are continuous, we discretise them by day or week or month, according to the number of observations we need per cluster. Let $T_{(i)}$ and $T^c_{(i)}$ be the times to next offer and of censoring of n independent patients. $\bar{T_i} = \mathbb{1}_{T_{(i)} \leq T^c_{(i)}} T_{(i)} + \mathbb{1}_{T_{(i)} \geq T^c_{(i)}} \mathbb{E}(T_{(i)} | T_{(i)} \geq T^c_{(i)})$ represents the observed time to next offer. The only assumption that we make here is that $\mathbb{E}(T_{(i)}) = \mathbb{E}(T_{(1)}),\ \forall i \in \{1, ..., n\}$ and $\mathrm{Var}(T_{(i)})$ does not "vary too much" with i. Thus, the distributions inside the cluster might be slightly different in terms of $T_{(i)}$ and even very different in terms of $T^c_{(i)}$. We have the following theorem:

Theorem 6.1.3 (Local mean with censored values)**.** *If* $\exists \epsilon < 1$*, and* $\exists m > 0$ *such that* $\forall i \in \{1, ..., n\}$*:*

(i) $\mathbb{E}(T_{(i)}) = \mathbb{E}(T_{(1)})$

(ii) $m\mathbb{E}\left(|T_{(i)} - \mathbb{E}(T_{(i)})|^3\right) \leq \mathrm{Var}(T_{(i)})$

(iii) $\mathbb{P}_{T_{(i)}, T^c_{(i)}}(T_{(i)} > T^c_{(i)}) \, \mathrm{Var}_{T_{(i)}, T^c_{(i)}}(T_{(i)} | T_{(i)} > T^c_{(i)}) \leq \epsilon \, \mathrm{Var}(T_{(i)})$

(iv) $\sum_{i=1}^n \mathbb{E}\left(|T_{(i)} - \mathbb{E}(T_{(i)})|^3\right) \underset{n \to +\infty}{\longrightarrow} +\infty$

Then:

$$n \frac{\tau_n - \mathbb{E}(T_{(1)})}{\sigma_n} \underset{n \to +\infty}{\xrightarrow{\mathcal{D}}} Z \sim \mathcal{N}(0, 1),$$

where: $\tau_n = \dfrac{\sum_{i=1}^n \bar{T_i}}{n}$ *and* $\sigma_n^2 = \sum_{i=1}^n \mathrm{Var}(\bar{T_i})$.

Remark 6.1.9. The theorem means that τ_n is an unbiased and consistent estimator of $\mathbb{E}(T_1)$ under some assumptions. The first assumption is that we are able to perfectly compute the best estimate of T under censorship ($\mathbb{E}(T_{(i)} | T_{(i)} > T^c_{(i)})$), which we do in practice by assuming that the estimated distribution is the actual distribution. Then, only mild assumptions on T and T^c are necessary. In our problem, they are always verified for T because it is exponential-like and because the distributions are quite similar to each other (due to the identity of the expected value).

Remark 6.1.10. We talked about clustering the observations by predicted expected time. If we group the observations by predicted month and then compute the local means at a day precision we are not consistent with our theorem. We should group by day and average by day. We claim that our result still holds when the predicted expected values are not assumed to be equal but in a same fixed interval. The proof is more technical and uses the fact that the Central Limit Theorem under Lyapunov's conditions allows for differences in the expected values.

Proof. The proof relies on the Central Limit Theorem under the conditions of Lyapunov.

Let us first find a lower bound for σ_n^2.

$$\begin{aligned}
\sigma_n^2 &= \sum_{i=1}^{n} \mathrm{Var}(\bar{T}_i) \\
&= \sum_{i=1}^{n} \mathrm{Var}(T_{(i)}) - \mathbb{P}_{T_{(i)},T^c_{(i)}}(T_{(i)} > T^c_{(i)}) \, \mathrm{Var}_{T_{(i)},T^c_{(i)}}(T_{(i)} | T_{(i)} > T^c_{(i)}), \text{ with theorem 6.1.2} \\
&\geq \sum_{i=1}^{n} \mathrm{Var}(T_{(i)}) - \epsilon \, \mathrm{Var}(T_{(i)}), \text{ by (ii)} \\
&\geq m(1-\epsilon) \sum_{i=1}^{n} \mathbb{E}\left(|T_{(i)} - \mathbb{E}(T_{(1)})|^3 \right), \text{ by (i) and (iii)}
\end{aligned}$$

Let us show that $L_n \underset{n \to +\infty}{\longrightarrow} 0$ with:

$$\begin{aligned}
L_n &= \frac{1}{\sigma_n^3} \sum_{i=1}^{n} \mathbb{E}\left(|\bar{T}_i - \mathbb{E}(T_{(1)})|^3 \right), \text{ with theorem 6.1.2 and hypothesis (i)} \\
&\leq \frac{\sum_{i=1}^{n} \mathbb{E}\left(|T_{(i)} - \mathbb{E}(T_{(1)})|^3 \right)}{m^{\frac{3}{2}}(1-\epsilon)^{\frac{3}{2}} \left(\sum_{i=1}^{n} \mathbb{E}\left(|T_{(i)} - \mathbb{E}(T_{(1)})|^3 \right) \right)^{\frac{3}{2}}}, \text{ with theorem 6.1.2 and previous inequality} \\
&= m^{-\frac{3}{2}}(1-\epsilon)^{-\frac{3}{2}} \left(\sum_{i=1}^{n} \mathbb{E}\left(|T_{(i)} - \mathbb{E}(T_{(1)})|^3 \right) \right)^{-\frac{1}{2}} \\
&\underset{n \to +\infty}{\longrightarrow} 0, \text{ by (iv)}
\end{aligned}$$

Thus we can use the Central Limit Theorem under the conditions of Lyapunov and derive the result from it.

□

From theorems 6.1.2 and 6.1.7, we can estimate confidence intervals for the local means, for n large enough. In practice:

Assumption 6.1.2. *For n large enough, under the null hypothesis, we have:*

$$\mathbb{E}(T^{\widetilde{\simeq}}_{(1)}) \in \left[\tau_n - 1.96\frac{\sqrt{\sum_{i=1}^n \mathrm{Var}(T^{\widetilde{\simeq}}_{(i)})}}{n}, \tau_n + 1.96\frac{\sqrt{\sum_{i=1}^n \mathrm{Var}(T^{\widetilde{\simeq}}_{(i)})}}{n}\right]$$

with 95% probability and $\tau_n = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{T^*_{(i)} \leq T^c_{(i)}} T^*_{(i)} + \mathbb{1}_{T^*_{(i)} > T^c_{(i)}} \mathbb{E}_{T^{\widetilde{\simeq}}_{(i)}}(T^{\widetilde{\simeq}}_{(i)} | T^{\widetilde{\simeq}}_{(i)} > T^c_{(i)})$.

6.1.8 Summary

Main Indicators

The above subsections go over several measures of performance for our algorithm. We introduced NSE, empirical mean quantiles and local means. The first one identifies something from the individual badness of a prediction, the second helps representing the overall quality of the underlying mathematical model and the third tackles the reliability of the predicted expected value. We give in table 6.1 a summary of all the measures which we mentioned with their particularities.

Remark 6.1.11. When explaining the C-index, we introduced the survival words calibration and discrimination (Pencina and D'Agostino, 2004). If the C-index is a discrimination measure, the empirical mean quantiles and the local means would be calibration measures in our problem. Discrimination and calibration measures are both absolute measures in our framework.

Complementary Indicators

In the line of studying the clinical performances of our algorithm (their meaning for the patient) we may want to look at other indicators such as accuracy and sensitivity.

- accuracy: $\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{T^*_{(i)} = \mathbb{E}(T^{\widetilde{\simeq}}_{(i)})}$. Due to the intrinsic stochasticity of the problem and clinical purposes, this only makes sense at a month granularity.
- sensitivity (first month prediction): $\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{T^*_{(i)} \leq 1} \mathbb{1}_{\mathbb{E}(T^{\widetilde{\simeq}}_{(i)}) \leq 1}$, at a month granularity.
- specificity (first month prediction): $\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{T^*_{(i)} > 1} \mathbb{1}_{\mathbb{E}(T^{\widetilde{\simeq}}_{(i)}) > 1}$, at a month granularity.
- confidence accuracy: $\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\mathbb{E}(T^{\widetilde{\simeq}}_{(i)})^- \leq T^*_{(i)} \leq \mathbb{E}(T^{\widetilde{\simeq}}_{(i)})^+}$, either at a day or month granularity.

Table 6.1 Validation measures: summary. The contributions of this work are in italic.

Measure	Use	Assumptions	Censoring	Comments
MSE	Relative		No	Straightforward to compare models
MAPE	Relative	Non-zero observations	No	Biased in favour of large observed times predicted small
SMAPE	Relative		No	Balanced between observed and predicted times
SAPE	Bad predictions		No	Balanced between observed and predicted times
MNSE	Relative	Null hypothesis	No	Biased in favour of small observed times predicted large
NSE	Bad Predictions		No	Biased in favour of small observed times predicted large
C-index	Relative, Absolute		Yes	Evaluates ranking power
1st MWD	Relative		No	More adapted to a training procedure than validation
2nd MWD	Relative		No	More adapted to a training procedure than validation
Empirical Mean Quantiles	Relative, Absolute	Independence	No	Qualitative and visual more than quantitative
Local Means	Absolute	Null Hypothesis, Independence, Mild Assumptions	Yes	Qualitative and visual more than quantitative

- conservativity: $\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{T^*_{(i)} \leq \mathbb{E}(T^{\simeq}_{(i)})}$, at a month granularity. We prefer to slightly overestimate than underestimate, not to give false hopes to the patient.

Censored Values

We proved only for a few indicators that we could include censored observations. We gave in theorem 6.1.2 an estimate of the time to next offer in a context of random censoring. Even without any theoretical support, we may want to use this estimator in the aforementioned indicators. Let us discuss the influence of the observed time estimator in case of a censored value. If the censoring time is very little and the predicted time is big, then we will have: $\bar{T}_i = \mathbb{E}(T_{(i)}|T_{(i)} > T^c_{(i)}) \simeq \mathbb{E}(T_{(i)})$ (slightly greater). In this case, the estimator is biased in favour of the model. However, if the censoring time is big (at least the same order of magnitude as $\mathbb{E}(T_{(i)})$), then the estimator is biased against the model $\bar{T}_i = \mathbb{E}(T_{(i)}|T_{(i)} > T^c_{(i)}) \gg \mathbb{E}(T_{(i)})$. Using this estimator is assuming that there is a worse bias in not taking censored values into account than in introducing this estimator. We should be careful if we use this estimator and we should not select models based on it. Instead, it provides an idea of what happens when we include censored values.

6.2 Results

We evaluate the performances of different variants of the algorithm predicting the time to next offer with different hyper-parameters.

6.2.1 Hyper-Parameters

For our algorithm, we have several hyper-parameters to choose. Each variant of the algorithm has different hyper-parameters (see section 4.3.2).

- Method: PWL or Current Waiting-List (CWL), with or without Eligibility Relaxation (ER).
  In practice, we could only test the ER with the CWL method.
- Number of days ΔT as a training set: 365 or 730.
- With or without cPRA.
  If we adapt the parameters μ_i to take into account the probability of positive cross-match for the current patient or not, by multiplying by $1 - \frac{\text{cPRA}}{100}$.
- Last rank of offer: historical rank or fixed rank.

- Number of bootstrap simulations n_{btp}.
  This needs a little tuning. We look for a trade-off between precision and computing-time.

We also mention hyper-parameters which we used as constants but which could be changed in further studies:

- Precision on the approximation of the expected value by early computation stopping: 0.1 day.

- Maximal cPRA: we took 99% for each model.
  A patient with a 100% cPRA will be considered having a probability 0.99 of being incompatible to a random donor. We apply this both for the patient for which we make predictions and the patients in the waiting-list for which we simulate compatibility in the CWL method.

- Number of waiting-list simulations in the CWL method n_{sim}: we took arbitrarily 1000 for each model.
  This is used to determine the minimal score of proposal corresponding to the last rank of proposal for each training donor. The blood-type eligible patients on the waiting-list are randomly assigned a cross-match (positive or negative) to the donor according to their cPRA. Only the patients with a higher score than the patient and a high cPRA are affected by this random draw. Under the assumption that the maximum incompatibility probability is 0.99, 1000 seems a reasonable choice.

Remark 6.2.1. We comment on the fact that we did not wish to tune hyper-parameters in a systematic way. We showed in previous section 6.1 that we did not have a single cost-function to evaluate our model. Thus we will base our judgement (of the different models and combinations of hyper-parameters) on subjective and qualitative considerations, summarising the largest number of factors. Consequently, it would not be consistent to automatically tune hyper-parameters to improve our validation performance on our specific validation set. This would directly lead to a (probably not serious in practice) over-fitting, considering the small amount of data.

6.2.2 Fixing some Hyper-Parameters

Number of Bootstrap Simulations

Some hyper-parameters are independent of each other. In particular, the number of bootstrap simulations only depends on the number of donors used in the training set. It only impacts the quality of the estimated confidence intervals, independently of the chosen method, because we are using the same distribution for T. This means that it is reasonable to tune n_{btp} on one of the configurations (e.g. the less time consuming for the highest number of training donors) before doing other experiments.

We computed the bootstrap intervals on the expected time to next offer for $n_{btp} \in \{1000, 10000, 20000, 30000\}$, $\Delta T = 730$, with cPRA for the Past Waiting-List method on our validation set. We give in table 6.2 the evolution of the absolute difference between the inferior and superior confidence bounds for two consecutive n_{btp} in the aforementioned set. As expected, the upper bound seems more stochastic than the lower bound, since there is a higher impact of removing an eligible donor from the training set than adding one on the expected time. Increasing n_{btp} improves the quality of the confidence intervals. There are still changes when increasing the parameter from 20000 to 30000 for less than half of the observations. We note that the gap remains big in terms of infinite norm. This is due to unexplained single values of donors, for which the upper confidence bounds seem to oscillate with the number of simulations.

Considering the computing time, which increases a lot with n_{sim}, we choose $n_{sim} = 10000$ in the rest of the study, towards validation. Of course, in practical use, n_{sim} can be increased, considering that there would be only one prediction to do at a time.

Number of Days for the Training Set

We only show for the less time-expensive method the impact of changing the number of days for the training set from 365 to 730. ΔT should be independent of the other hyper-parameters, because it has only two effects: a smaller ΔT makes the model more sensitive to changes in the distribution of donors, a larger ΔT provides more donors to estimate the distribution.

Some relative indicators of performance (needing no assumption) are presented in table 6.3 in terms of prediction of expected time to next offer. Even though some indicators are

Table 6.2 Different distances between 95% bootstrap confidence bounds for different numbers of simulations for the validation set including censored values. The three quantiles of the absolute difference between bounds are given too. Method Past Waiting-List with cPRA and $\Delta T = 730$.

$\lvert T(n_{sim}) - T(n'_{sim})\rvert$	$\lvert\lvert\cdot\rvert\rvert_1$	$\lvert\lvert\cdot\rvert\rvert_2$	$\lvert\lvert\cdot\rvert\rvert_\infty$	25%	50 %	75%
|1000 - 10000|: lower	0.69	1.86	18.78	0.0	0.03	0.52
|1000 - 10000|: upper	3.42	10.57	157.23	0.0	0.24	1.97
|10000 - 20000|: lower	0.18	0.7	8.15	0.0	0.0	0.04
|10000 - 20000|: upper	1.57	5.62	80.68	0.0	0.01	0.57
|20000 - 30000|: lower	0.17	0.85	12.32	0.0	0.0	0.03
|20000 - 30000|: upper	0.98	3.56	41.5	0.0	0.01	0.31

in the same order of magnitude, the MSE is more than one third lower for $\Delta T = 730$ than $\Delta T = 365$. Even the predictive power is better with a larger time window, as shown by the little increase in C-index (which is significant in our opinion, considering the fact that the C-index is computed over all possible pairs of observations: for the dataset, an order of magnitude of 10^5 pairs). For MAPE and 2nd MWD however, the results are slightly worse for $\Delta T = 730$. This is due to the fact that the results were probably underestimated for $\Delta T = 365$ for some patients (as a result of a lower number of training donors, leading to less granularity in the high predicted times), in comparison to $\Delta T = 730$ where the results were more likely overestimated, leading to higher predicted variances penalised by the 2nd MWD.

This analysis leads us naturally to choose $\Delta T = 730$.

Table 6.3 Comparison of expected time prediction performances between $\Delta T = 365$ and $\Delta T = 730$. The Past Waiting-List method with cPRA was used on the validation set (including censored data for the C-index).

ΔT	MSE	MAPE	SMAPE	1st MWD	2nd MWD	C-index
730	7525.2	580.51	85.01	64.07	82.81	0.7229
365	12704.79	568.78	85.15	66.75	79.1	0.7167

6.2.3 Validation

Comparison of Numerical Indicators

We give in table 6.4 our numerical validation results. We can infer different things from it.

We first comment on the impact of adjusting for the positive-crossmatch probability with the cPRA. It always worsens the prediction in terms of MSE, MAPE, SMAPE and MWD. This means that we are over-adjusting for positive-crossmatch probability. This brings evidence that the cPRA as used today is not well calibrated as a probability of incompatibility among a random set of donors (or that we neglected latent interactions in the model). However, we also note that the difference is less dramatic for the Past Waiting-List method than for any configuration of the Current Waiting-List method. The fact that the converse phenomenon is observed for MNSE supports our claim that this indicator can only be used for identification of bad predictions under the null hypothesis as it favours over-predictions. Very interestingly, the C-index shows that taking cPRA into account improves the predictive power of the method (of at least 0.015 for each method). This improvement is significant in our opinion, considering the fact that the C-index is computed over all possible pairs of observations as observed before. This supports the fact that cPRA has an impact on probability of tissue-type incompatibility, but is not linear with the cPRA.

We comment on the Last Rank of offer. We only provided experiments with Last Rank set at 3 for the Current Waiting-List method without Eligibility Relaxation. The results are at least 10 times as bad as any other method in terms of MSE and they are worse for any other relevant relative indicator, including C-index. This brings evidence that the historical rank of last offer is a relevant measure of donor popularity. Therefore, we only keep the historical rank or past rank of last offer for our method.

We comment on Eligibility Relaxation. ER improves the MSE, whereas MAPE, SMAPE, 1st and 2nd MWD do not show any significant improvement or worsening in our opinion. However it seems to worsen accuracy, confidence accuracy and first month sensitivity. First month specificity remains the same as well as conservativity. These parameters are hard to interpret here, because it is obvious that the CWL method greatly over-estimates all the predictions. The difference between ER and no ER seems more subtle than a systematic over or under-estimation. However, there is a slight improvement in terms of C-index, which in our opinion is significant, and, in addition to MSE, supports our mathematical intuition in section 4.2.3.

Table 6.4 Numerical validation results for several configurations of hyper-parameters. For each row, the maximum is in bold and the minimum in italic.

Method	CWL	CWL	CWL	CWL	CWL	CWL	PWL	PWL
ER	No	No	No	No	Yes	Yes	No	No
Last Rank	3	3	Past	Past	Past	Past	Past	Past
ΔT	730	730	730	730	730	730	730	730
cPRA	No	Yes	No	Yes	No	Yes	No	Yes
Maximal cPRA	99	99	99	99	99	99	/	99
n_{btp}	10000	10000	10000	10000	10000	10000	10000	10000
n_{sim}	1000	1000	1000	1000	1000	1000	/	/
MSE	237078	**252116**	25325	29224	21365	25491	*7051*	7525
MAPE	2889	**3279**	736	984	731	1029	*391*	581
SMAPE	130.5	**133.1**	95.3	98.9	95.3	101.0	85.6	*85.0*
MNSE	4.5	*0.28*	4.85	1.08	3.06	1.17	**7.16**	1.86
1st MWD	313.7	**332.5**	95.0	108.1	92.2	107.5	*57.6*	64.1
2nd MWD	228.6	**251.6**	109.7	128.2	115.5	138.2	*71.4*	82.8
Accuracy (months)	0.24	*0.21*	0.37	0.3	0.36	0.27	**0.43**	0.38
Confidence Accuracy	0.05	*0.04*	0.2	0.17	0.16	0.13	**0.24**	0.24
Confidence Accuracy (months)	0.31	*0.26*	0.56	0.5	0.51	0.43	**0.67**	0.63
Conservativity (months)	0.94	**0.98**	0.86	0.92	0.87	0.92	*0.74*	0.8
Sensitivity (1st month)	0.4	*0.33*	0.52	0.4	0.5	0.36	**0.62**	0.52
Specificity (1st month)	0.91	**0.98**	0.84	0.95	0.86	0.95	*0.77*	0.88
C-index	*0.68*	0.6921	0.7122	0.732	0.7182	**0.7379**	0.7039	0.7229

Finally, we comment on the performances of the Past Waiting-List method against Current Waiting-List. The former beats the latter in terms of MSE, MAPE, SMAPE, MWD, accuracy, confidence accuracy, sensitivity. However, as the CWL over-estimates the expected time to next offer, it is of course more specific and conservative than PWL. Again, the C-index brings an interesting perspective on both methods. The C-index is 0.02 lower than the best performing CWL. This supports the impact of the actual current waiting-list of the patient in the distribution of eligible donors. Furthermore, together with previous statement, this supports the fact that using the historical last rank of offer is relevant for the current waiting-list. This means the CWL method is more promising but not calibrated. This calibration problem originates in the identification of the eligible donors. This is due to any of the following problems:

- We do not retrieve correctly the current waiting-list, and we include patients who should not be active and unduly score a lot in waiting-time.
- We do not correctly take into account priority lists and systematically include priority patients in the general attribution list. Considering the complex attribution system, we should account for priority patients before including them in the general attribution list.
- We do not correctly simulate the tissue-type compatibility in the current waiting-list for the potential donors, which is supported by our previous observation about cPRA.

We comment on the absolute performances of the PWL method. The accuracy in months is between 40% and 50%. The confidence accuracy in days is of course very low (20%) and the confidence accuracy in months is between 60% and 70%. There is no reason for these indicators to be good *a priori*. Indeed, they measure the ability of the predicted expected time to predict the actual time to next offer. The confidence bounds are no confidence bounds on the actual time to next offer but on the quality of the estimation of the expected value. So those indicators measure the clinical relevance of giving the predicted expected time to next offer to the patient. In our opinion, those results support the choice of the expected value as a time to give to the patient. One would ask why we do not give the median time instead. In our model, the expected time is higher than the median time. As we prefer to over-estimate than under-estimate the predicted time, not to give false hopes to the patients, the expected time is more suitable. In months, the predicted time is higher than the observed time in 70% to 80% percent of the observations. About the ability to predict an offer in the first month, without cPRA, the model is 60% sensitive and 70% specific. With cPRA, it is 50% sensitive

and 85% specific. In general, we are better in predicting less than one month than more than one month.

Visualisation of Performance

The numerical indicators enabled us to evaluate the relative quality of the models and the absolute quality of the predicted expected times. Now, we propose to evaluate the global quality of the models and the potential of using quantiles in the prediction. Considering the poor performances provided by Last Rank set to 3, we remove it from the following studies.

We give in figures 6.1 and 6.2 the local means with confidence intervals as defined in section 6.1.7. Firstly, we comment on the upper-bound for predicted expected times in months: we kept only the means for which we had at least 15 observations. This led us to a maximum predicted 5 months for the uncensored case and 8 months including censored value. To let the reader know, the means get very chaotic for the uncensored case after 5 months. For the censored case, it got chaotic after the 8th month. Yet, for some isolated values, the predicted mean were deceptively accurate, due to censoring after a very short time, leading to $\bar{T}_i = \mathbb{E}(T_{(i)}|T_{(i)} > T^c_{(i)}) \simeq \mathbb{E}(T_{(i)})$.

For both figure 6.1 and 6.2, the CWL method ranks the different means almost in the correct order, even though the predictions are over-estimated (remember that the plot displays the average of the observed values). For the PWL method with cPRA, excluding censored values gives good predictions up to the 4th month, whereas including them gives convincing predictions up to the 5th month, by visual inspection of the curves. All the algorithms give very accurate means for the first month, which is consistent with previous results and with the model, as stochasticity increases with the predicted time. There is no significant change in taking cPRA into account or not, apart for the first month (even though the average in months is still respected) and the second month (which gets out of range when neglecting cPRA).

From this analysis, it results that the CWL method has a good potential, consistent with the high C-index, but the PWL is quantitatively better.

Remark 6.2.2. We chose to cluster the observations by predicted month rather than any other choice to be consistent with our numerical verifications. One may argue that we could have adapted the clusters to always have the same number of observations per cluster. We thought it was not a good idea to enlarge the prediction window (e.g. 6 months to 12 months)

because then we were very unsure if our results still held regarding the conditions of theorem 6.1.3. Conversely, we did not wish to shrink those intervals as we got near 0 predicted month, because we do not use more than a month accuracy in clinical practice.

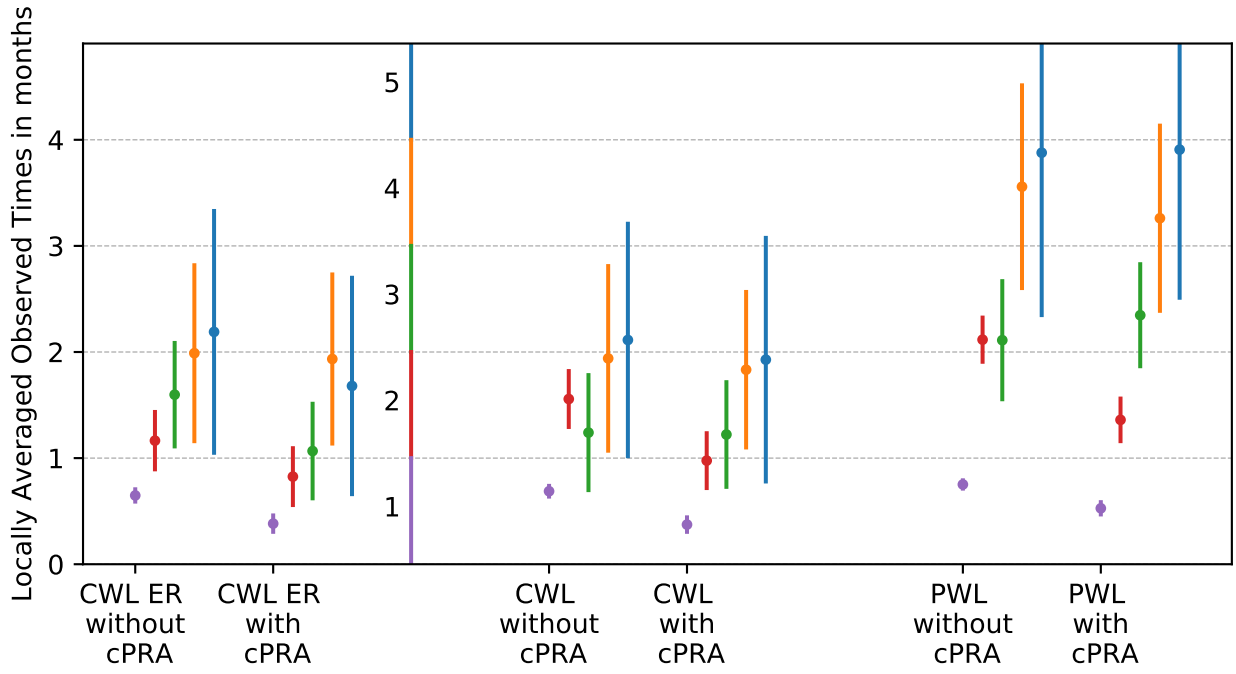

Figure 6.1 Local means for different methods with expected confidence intervals excluding censored data from the validation set. There are at least 15 observations per mean.

We provide in figure 6.3 the empirical mean quantiles with confidence bounds against the expected quantiles. The values are given in table 6.5. Again, the CWL method is very conservative but overestimates all the quantiles. On the opposite, the PWL method provides coherent quantiles over the whole window. However, accounting for cPRA gives better high quantiles and neglecting it gives better low quantiles. Our preference however is for including cPRA, as the very low quantiles are still convincing and the high quantiles very accurate, meaning we are covering the full range with our method, neither underestimating or overestimating in general.

We give in figure 6.4 the empirical mean quantiles including censored data, with the method explained in section 6.1.8. Indeed, even though there is no theoretical support for doing so, the results turned out to look very good, with an almost perfect fit for each quantile for

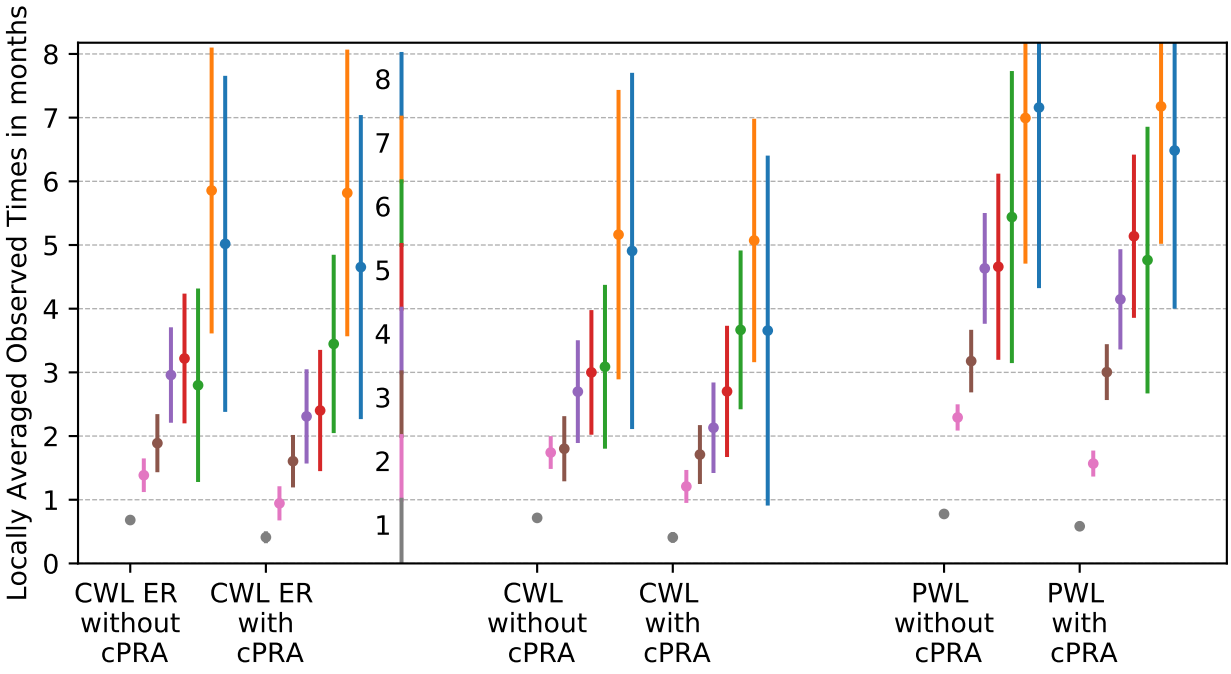

Figure 6.2 Local means for different methods with expected confidence intervals including censored data in the validation set. There are at least 15 observations per mean.

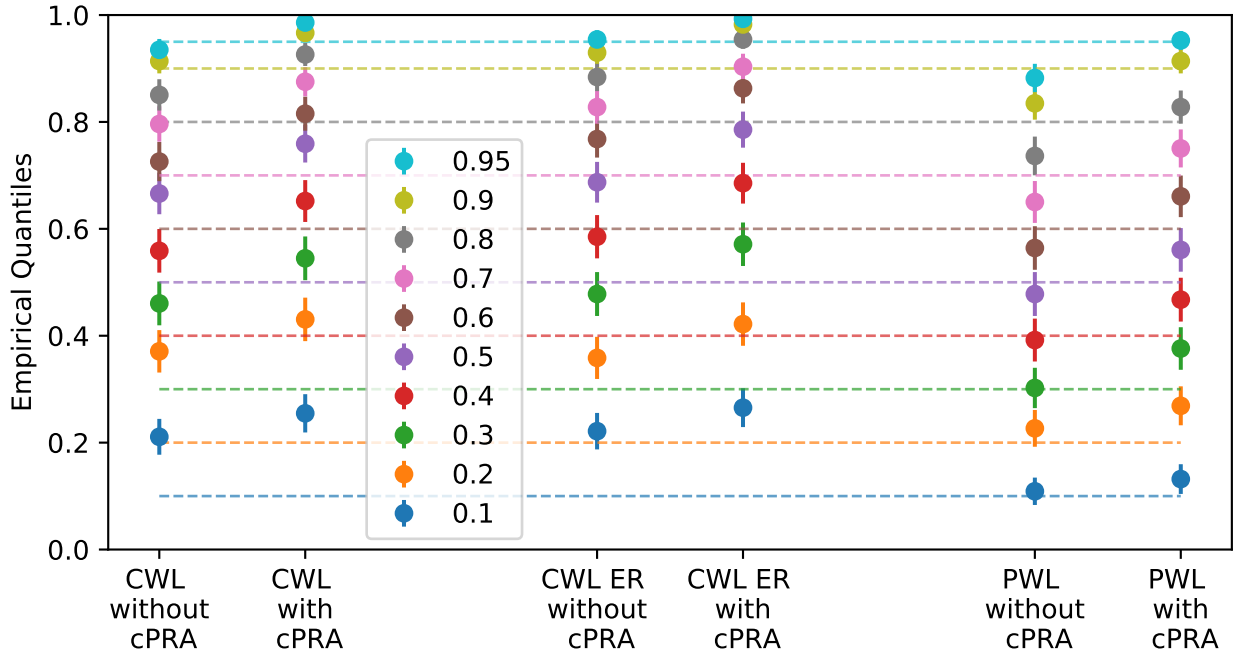

Figure 6.3 Empirical mean quantiles for different methods with confidence intervals excluding censored data from the validation set.

Table 6.5 Values of the empirical mean quantiles with confidence intervals for different sets of hyper-parameters. We removed the censored values from the validation set. The closest values to the expected ones are in bold.

Method	CWL	CWL	CWL	CWL	PWL	PWL
ER	Yes	Yes	No	No	No	No
Last Rank	Past	Past	Past	Past	Past	Past
ΔT	730	730	730	730	730	730
cPRA	No	Yes	No	Yes	No	Yes
Maximal cPRA	99	99	99	99	/	99
n_{btp}	10000	10000	10000	10000	10000	10000
n_{sim}	1000	1000	1000	1000	/	/
0.1	$_{0.187}0.221_{0.256}$	$_{0.229}0.265_{0.302}$	$_{0.177}0.211_{0.244}$	$_{0.219}0.255_{0.291}$	$_{0.083}\mathbf{0.109}_{0.135}$	$_{0.104}0.132_{0.16}$
0.2	$_{0.319}0.359_{0.398}$	$_{0.381}0.422_{0.462}$	$_{0.331}0.371_{0.411}$	$_{0.39}0.431_{0.471}$	$_{0.192}\mathbf{0.227}_{0.261}$	$_{0.232}0.269_{0.305}$
0.3	$_{0.437}0.478_{0.519}$	$_{0.531}0.571_{0.612}$	$_{0.42}0.46_{0.501}$	$_{0.504}0.545_{0.586}$	$_{0.265}\mathbf{0.302}_{0.34}$	$_{0.336}0.376_{0.416}$
0.4	$_{0.545}0.585_{0.626}$	$_{0.647}0.685_{0.724}$	$_{0.518}0.559_{0.6}$	$_{0.613}0.652_{0.691}$	$_{0.352}\mathbf{0.392}_{0.432}$	$_{0.426}0.467_{0.508}$
0.5	$_{0.649}0.687_{0.725}$	$_{0.752}0.786_{0.819}$	$_{0.627}0.666_{0.705}$	$_{0.724}0.759_{0.794}$	$_{0.437}\mathbf{0.478}_{0.519}$	$_{0.52}0.561_{0.601}$
0.6	$_{0.733}0.768_{0.803}$	$_{0.835}0.863_{0.891}$	$_{0.689}0.726_{0.762}$	$_{0.784}0.815_{0.847}$	$_{0.523}\mathbf{0.564}_{0.605}$	$_{0.622}0.661_{0.7}$
0.7	$_{0.797}0.828_{0.859}$	$_{0.879}0.903_{0.928}$	$_{0.763}0.796_{0.829}$	$_{0.848}0.875_{0.902}$	$_{0.611}\mathbf{0.65}_{0.689}$	$_{0.715}0.75_{0.786}$
0.8	$_{0.858}0.884_{0.91}$	$_{0.937}0.954_{0.971}$	$_{0.821}0.851_{0.88}$	$_{0.905}0.926_{0.948}$	$_{0.7}0.736_{0.773}$	$_{0.797}\mathbf{0.828}_{0.859}$
0.9	$_{0.909}0.93_{0.951}$	$_{0.972}0.982_{0.993}$	$_{0.891}\mathbf{0.914}_{0.937}$	$_{0.952}0.967_{0.981}$	$_{0.804}0.835_{0.865}$	$_{0.891}0.914_{0.937}$
0.95	$_{0.937}0.954_{0.971}$	$_{0.986}0.993_{1.0}$	$_{0.915}0.935_{0.955}$	$_{0.976}0.986_{0.996}$	$_{0.856}0.882_{0.909}$	$_{0.935}\mathbf{0.953}_{0.97}$

method PWL with cPRA. In presenting this figure, we make the assumption that the pros in including censored values outweigh the cons of using an estimator with no theoretical support.

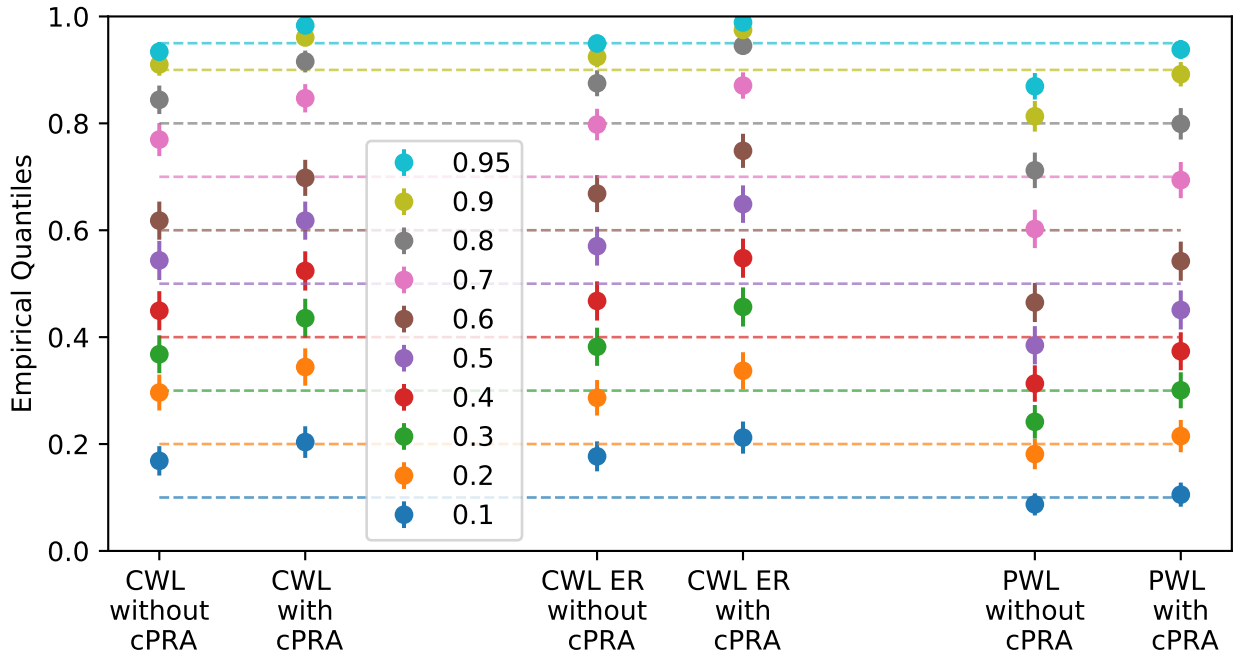

Figure 6.4 Empirical mean quantiles for different methods with confidence intervals including censored data in the validation set. Theorem 6.1.2 was used to retrieve observed times.

For our best method, the predicted $t_{95\%}$ is very reliable (unlike the predicted median). This supports the idea that we should give to the patient both predicted expected time $\mathbb{E}(T)$ and the time $t_{95\%}$ for the offer, as both these values our reliable, complementary and understandable.

Identification of Bad Predictions

We present below the pragmatic method which we used several times in our work to detect bad predictions (as defined in section 6.1.3). This helped us to detect mistakes in score calculation or in the data. In figures 6.5 and 6.6, we calculated the Symmetric Absolute Percentage Error (SAPE) (resp. NSE) for each observation in the validation set (excluding censored data) and for different experiments. Then we sorted these values in increasing order for each experiment and plotted them.

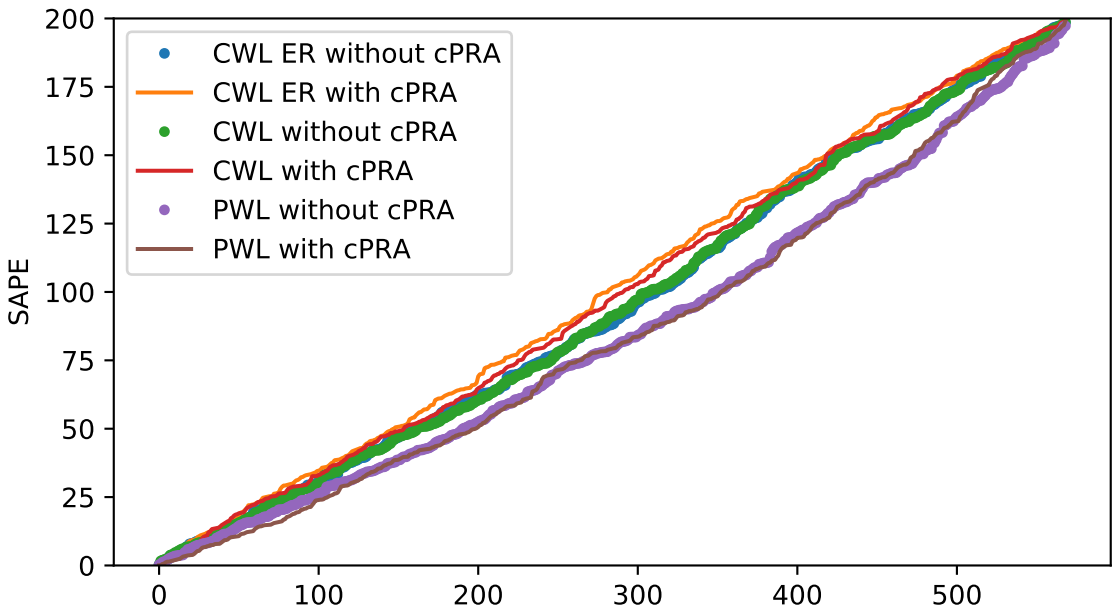

Figure 6.5 Symmetric Absolute Percentage Error between observed and predicted expected time in increasing order for different experiments. We excluded censored values from the validation set.

The SAPE (see figure 6.5) is interesting because we can confirm that the PWL method performs better than the others. However, this measure does not enable an obvious split between good and bad predictions in our case. This inherent to the fact that the indicator is bounded and does not include a squared error.

Our NSE (see figure 6.6) looks more adapted but should be used with care. Indeed, looking at the figure, one would think the CWL methods work better than PWL. This is forgetting that this indicator is good at identifying bad predictions the closer we are to the null assumption, which is somehow paradoxical. A great value of NSE implies necessarily a bad prediction but the opposite is not true, because it favours over-estimations. Therefore, we will focus on the PWL as we consider that it is our best method, but we can use the other methods to check if all methods predicted badly the same observations. Looking at the figure, it seems that a good threshold for "bad" predictions is a NSE above 3.

We compared the bad predictions for the PWL method with and without cPRA adjustment. It shows the necessity of taking cPRA into account, as extremely high values of NSE

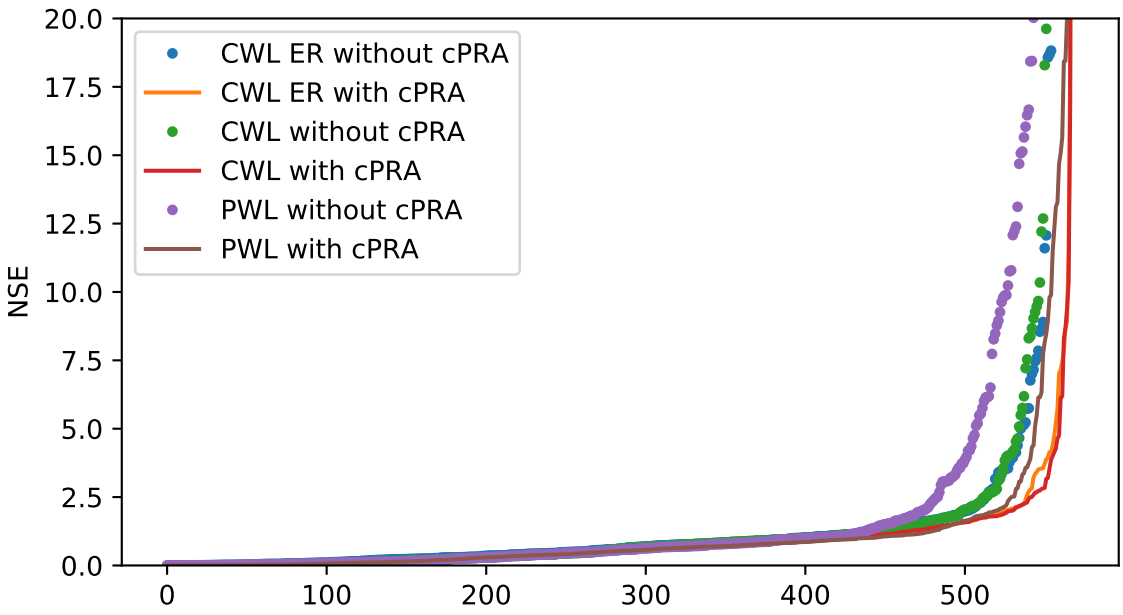

Figure 6.6 Normalised Squared Error between observed and predicted expected time in increasing order for different experiments. We excluded censored values from the validation set. For display purposes, we set an upper-limit of 20 for the y-axis.

are avoided through the adjustment. The adjustment will usually only reduce the NSE so we can focus on the results without cPRA. We study the difference between the population of all the validation samples and the validation samples with NSE greater than 3. We excluded censored values. Statistics are presented in tables 6.6 and 6.7. We note that, for the whole validation set, the mean of predicted expected times is almost the same as the mean of the observed times. This is probably not due to chance, but more likely to the Central Limit Theorem, showing that our algorithm has good calibration properties. As expected, there are far more sensitised patients among bad predictions, with level of sensitisation between 40% and 97%. This brings evidence that the cPRA is not well calibrated for medium to high cPRA (under-estimating the probability of incompatibility), but not for very high cPRA (either well calibrated or over-estimated). However, at least 25% of bad predictions occur for non-sensitised patients. This is harder to explain. It might be caused by our 0/1 eligibility system (then eligibility relaxation would solve the problem), too little data (sequence of unlikely good matches in the training set) or to latent variables. At the same time, we observe that the predicted times are under-estimated for bad predictions (obviously because NSE is meant to detect under-estimations), but the extreme values show that bad predictions do

not necessarily occur for large observed times.

We also looked at blood-type distributions. However, we did not see any remarkable change in the blood-type distributions, between bad predictions and all predictions together. If it had been different, and especially, if the proportion of AB patients had been higher for bad predictions, it would have brought evidence that the number of donors in the training set impacts the results. Indeed, AB is the rarest blood-type among patients and donors, leading to less occurrences of AB donors in the training set and reducing the size of the training set.

To put bad predictions in perspective and measure their importance, we also compare the proportion of bad predictions with and without cPRA adjustment in table 6.8. This supports the fact that the larger errors are due to cPRA.

Computational Times

The time of computation depends on the hyper-parameters. We give in table 6.9 an idea of the computational times for different sets of hyper-parameters. We only included the hyper-parameters which are supposed to have an impact on the computing time. There is a small difference when taking cPRA into account which we did not choose to present here. One may wonder why there is such a difference between the PWL and CWL methods. This is due to the way we handle data and to the fact that we have to infer the historical (current) waiting-list for the CWL method for each observation. For the PWL method, we did not have to retrieve the waiting-list at each historical date, because we could directly infer the score of last offer from the data. That explains however why we did not use the ER on the PWL: we would have had to know the rank of the patient in each historical waiting-list, to be able to compute the eligibility probability.

Table 6.6 Statistics about time to next offer and cPRA in the validation set excluding censored values for predictions with NSE higher than 3. PWL method without cPRA. This corresponds to 72 different patients.

Feature	Count	Mean	Std	Min	25%	50 %	75%	Max
T^*	83.0	148.69	146.43	17.0	36.0	87.0	220.0	667.0
$\mathbb{E}(T^{\simeq})$	83.0	34.53	54.51	5.18	6.43	16.16	45.3	445.98
cPRA	83.0	41.04	35.37	0.0	0.0	43.0	72.5	97.0

Table 6.7 Statistics about time to next offer and cPRA in the validation set excluding censored values for predictions. PWL method without cPRA. This corresponds to 325 different patients.

Feature	Count	Mean	Std	Min	25%	50 %	75%	Max
T^*	569.0	54.56	84.5	0.0	8.0	24.0	63.0	667.0
$\mathbb{E}(T^{\simeq})$	569.0	55.22	67.21	5.18	11.9	34.76	66.42	535.79
cPRA	569.0	19.44	26.66	0.0	0.0	2.0	30.0	99.0

Table 6.8 Proportion of bad predictions with NSE in the validation set for the PWL method with and without cPRA adjustment.

Proportion in the validation set	NSE ≥ 3	NSE ≥ 10
PWL without cPRA	14.6%	7.4%
PWL with cPRA	6.2%	2.5%

Table 6.9 Prediction of one distribution of next offer: average computational times with standard deviation for different sets of hyper-parameters.

Method	CWL	CWL	PWL	PWL	PWL	PWL	PWL
ER	Yes	No	No	No	No	No	No
ΔT	730	730	365	730	730	730	730
n_{btp}	10000	10000	10000	10000	1000	20000	30000
n_{sim}	1000	1000	/	/	/	/	/
Mean Time (std)	50s (10)	44s (7)	12s (6)	14s (6)	5s (0.5)	26s (12)	38s (18)

Summary

From this validation study, it is difficult to draw categorical choices. The Current Waiting-List method with Eligibility Relaxation seems the most promising method, due to its discriminative power, though it is not calibrated. The Past Waiting-List method is well calibrated and has a good discriminative power, although not as good as the other method.

We make the pragmatic choice of the Past Waiting-List method with cPRA, $\Delta T = 730$, $n_{btp} = 10000$ in practice.

6.2.4 Test

Considering the fact that we did not actually tune our hyper-parameters here, the testing procedure will not highlight over-tuning but variability in the results. Nevertheless, we give a few results on the test set for our Past Waiting-List method with cPRA, which we compare to the validation set.

We give in table 6.10 numerical indicators of performance as well as the empirical mean quantiles for test vs. validation set. As we expected, the variation for the empirical mean quantiles is not significant, i.e. stays within the estimated confidence intervals. Our algorithm is slightly more sensitive to the first month on the test set than on the validation set, even though this is probably not significant considering the variability of the value (near 0.5). The C-index however is 0.02 greater on the test set than on the validation set. This is interesting and supports the idea that our algorithm performs better on some types of patients than on others (which would be caused by the variability of the distribution when splitting the datasets).

We give the local means for both uncensored and censored data in figure 6.7. Once more, even if there seems to be changes for some means, those are not significant with respect to the confidence intervals. This supports the fact that our prediction (in terms of expected time to next offer) is relevant on average up to the fourth month.

6.2.5 Summary

After the validation procedure, we considered that the PWL method was performing best and we evaluated it on the test set. The results show that the algorithm predicts faithfully the distribution of time to next better offer, excepted for highly-sensitised patients and some rare

Table 6.10 Numerical validation vs. test results for the PWL with cPRA, $\Delta T = 730$, $n_{btp} = 10000$. For each row, the maximum is in bold, excepted for the empirical mean quantiles for which the closest value to the expected one is in bold.

	Test	Validation
Accuracy (months)	**0.39**	0.38
Confidence Accuracy	0.21	**0.24**
Confidence Accuracy (months)	0.63	**0.63**
Conservativity (months)	0.79	**0.8**
Sensitivity (1st month)	**0.56**	0.52
Specificity (1st month)	0.86	**0.88**
C-index	**0.744**	0.7229
0.1	$_{0.11}0.139_{0.167}$	$_{0.104}\mathbf{0.132}_{0.16}$
0.2	$_{0.207}\mathbf{0.243}_{0.278}$	$_{0.232}0.269_{0.305}$
0.3	$_{0.335}\mathbf{0.374}_{0.414}$	$_{0.336}0.376_{0.416}$
0.4	$_{0.448}0.489_{0.53}$	$_{0.426}\mathbf{0.467}_{0.508}$
0.5	$_{0.536}0.576_{0.617}$	$_{0.52}\mathbf{0.561}_{0.601}$
0.6	$_{0.633}0.671_{0.71}$	$_{0.622}\mathbf{0.661}_{0.7}$
0.7	$_{0.713}\mathbf{0.749}_{0.784}$	$_{0.715}0.75_{0.786}$
0.8	$_{0.802}0.833_{0.864}$	$_{0.797}\mathbf{0.828}_{0.859}$
0.9	$_{0.875}\mathbf{0.9}_{0.924}$	$_{0.891}0.914_{0.937}$
0.95	$_{0.921}0.94_{0.96}$	$_{0.935}\mathbf{0.953}_{0.97}$

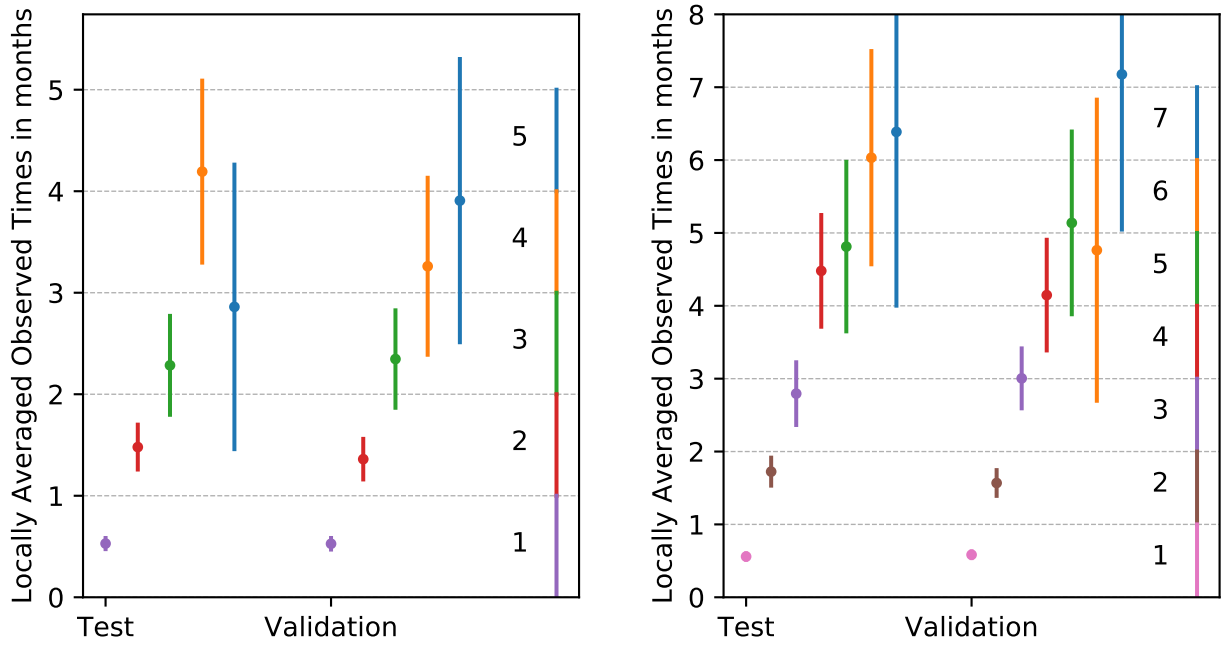

Figure 6.7 Local means for the validation and test set on the PWL method with cPRA. Excluding censored values (left) and including censored values (right).

cases. This procedure only evaluated the ability of the algorithm to predict the distribution of time to next offer. However, as shown in section 4.3, we can include an estimation of the quality of the next offer (e.g. through the KDRI) and an estimation of the time to next better offer (with also an estimation of the quality).

CHAPTER 7 CONCLUSION

7.1 Summary

We address the problem of informing an ESKD patient on his perspectives about a next offer if he rejects the current kidney offer. We restricted our study to non-paediatric patients on a general scoring waiting-list without special priorities.

This problem is very important in the decision process both for patient and physician. Having information only about the expected survival for the current kidney offer is incomplete. Information is needed on the expected survival relatively to the other potential offers. Information is also needed on the expected supplementary waiting-time for such an offer. This is difficult because of the many layers in the attribution process and the many requirements for a patient to first get an offer and then be possibly transplanted.

We modelled the arrival of eligible donors for a patient as a non-homogeneous Poisson point process with piecewise constant arrival rate. This enabled us to model the distribution of time to next offer. The method used to estimate the parameters gave us a representation of the distribution of the eligible donors. Thus, we were able to give an estimation of the quality of the next offer for any given measure of quality, and estimate the distribution of next better offer for this quality measure. To address the specific structure of the problem, we developed new indicators to validate our algorithm on real world data from Québec. Our best performing variant of the algorithm showed consistent results to predict the distribution of time to next offer up to four months. Certain variants of the algorithm showed very promising results even though they were not well calibrated.

Our main contribution is on the decision-aid methodology. We think that probabilities are difficult to interpret by patients and even by educated specialists. Therefore, giving an information in terms of time is more appropriate in our opinion. Our solution is disruptive in the way of empowering patient and physician in the decision-making. Indeed, we think that informing a patient is not about providing an optimal decision assuming his preferences. We also think that, for a personalised decision-support tool, personalised information should be provided. Therefore, our second contribution is on the validation methodology, including attempts to evaluate the quality of the whole model as well as the quality of single predictions.

7.1.1 Two Practical Examples

As a summary of the practical use of our work, we propose to give two examples of the practical use of our algorithm on real-world examples.

A "Low-Priority" Patient

We consider a 35 years old female non-sensitised AB transplant candidate. She already spent three months on dialysis. She gets an offer for a 53 years old AB Caucasian donor, with BMI 25. We mention that this donor died from a stroke (DND donor), and was an active smoker. The full KDPI for this donor is 73% (meaning that this kidney is worse than 73% of the kidneys of the year before the offer).

For this offer, we predict 221 days of expected time to next offer (7 months, real value between confidence bounds 139 days and 524 days), before 20 months with 95% confidence, and a KDPI of 66%. The estimation was made with an average of 3.4 eligible donors (the distribution of eligible donors vary with time). Considering the little number of eligible donors used to make the prediction, we do not compute the time to next better offer. The actual time to next offer was 104 days with a KDPI 57%.

The patient could have been informed accordingly: "We expect that on average, you will need to wait for another 7 months, and most certainly less than 20 months. Among last year's donors, 73% are expected to provide longer survival than the current one. We estimate that the average quality of the donor for the next offer for you will be slightly better than this one, as this time 66% of donors in the last year would be expected to provide longer survival."

A "High-Priority" Patient

We consider a 73 years old male patient with blood-type A and a cPRA of 26%. He has been under dialysis for 16 months. He gets an offer for a 63 years old male Caucasian A donor, with BMI 25. The donor died from a stroke (DND donor), had hypertension and was an active smoker. The full KDPI for this donor was 100%.

We predict an expected time to next offer of 18 days (real value between confidence bounds 15 and 23 days), before 8 weeks with 95% confidence, and a KDPI 97%. The estimation was made with an average of 54 eligible donors. We also predict a time to next better offer of 25 days (real value between confidence bounds 20 and 33 days), before 10 weeks with 95%

confidence, and a KDPI 92%. The estimation was made with an average of 40 eligible donors. The actual time to next offer was 12 days with a KDPI 97%.

The patient could have been informed accordingly: “We expect that on average, you will need to wait for 2 or 3 weeks and and most certainly less than 8 weeks. Hundred percent of last year’s donors are expected to provide longer survival than the donor who is currently proposed to you. On average, 97% of last year’s donors are expected to provide longer survival than the next offer. If you want to have a better offer, you will wait in average 3 or 4 weeks and most certainly less than 10 weeks. This time, 92% of last year’s donors are expected to provide longer survival than the next better offer.”

Moral

The doctor should evaluate if it is appropriate to ask for a prediction of next better offer. It is possible to ask the patient for his minimal expectations in terms of quality and try to predict the waiting-time necessary to achieve this quality. The KDPI is difficult to interpret and to communicate faithfully to the patient because of lack of personalisation and of significance, even though it is relevant in terms of comparison with the pool of donors.

7.2 Limitations

Our model has some limitations and our verification procedure has other limitations.

Our algorithm is inherently poor at predicting distributions of next offer of low-priority or hard-to-match patients. To predict the distribution of next offer for a low-priority patient, we have little eligible donors and thus we have a sparse representation of the set of eligible donors and an inaccurate time to next offer. This inaccuracy adds to the natural stochasticity of offers for low-priority patients. For hard-to-match patients, the next offer is heavily dependent on isolated very good donors which are hard to forecast. Furthermore, the level of sensitisation as represented by the cPRA is not mathematically reliable, so that we do not take sensitisation accurately into account.

Another limitation is the verification. We tried to remove all observations with mistakes. Yet, this may have biased our set and some mistakes may have remained. For example, if we removed patients with wrong dates of first dialysis and that those patients are long-waiting patients on waiting-list with high scores, we introduced a bias in the validation and test

sets as well as in the training set for some of the variants of the algorithm. This could be addressed by doing further experiments on synthetic data.

The two major limitations of the work for practical use concern the quality of a patient-donor match and the death or removal from waiting-list hazard. We do not have so far a reliable enough quality measure to fit in our information pedagogy: there is no model predicting a time of survival with confidence bounds to give times to the patient instead of probabilities. Additionally, we do not know the probability that a patient dies on the waiting-list or gets temporarily or permanently removed from it. We entrust to the physician the evaluation of the hazard, but we did not include it in our information framework.

7.3 Future Research Directions

Future research directions include addressing the aforementioned limitations. Beyond these, there is space for theoretical and experimental improvements of the solution proposed, as well as other research topics with the same dataset.

We can think about several theoretical improvements. In this work, we used donors eligible to a patient to estimate the expected quality of the next kidney offer. However, this does not enable a straightforward estimation of the confidence intervals on the expected quality and does not give a very granular representation of the distribution of eligible donors, particularly in a situation of sparse data. Therefore, we think that including priors, as in a Bayesian framework, might help representing this distribution and make full use of the available data. This could be an idea to increase granularity and robustness in the time prediction for low-priority patients too. Besides, we introduced new indicators to verify our algorithm, but in order to use them with more confidence, it would be valuable to run simulations to assess the sensitivity and the specificity of these indicators. Moreover, we saw that the version of the algorithm using the current waiting-list as a test list for donor eligibility was promising but not calibrated. Thus, we think that complementary analyses to better take into account patients on priority lists could contribute solving the problem. Finally, we would like to verify our predictions of expected time to next better offer. A mathematical work is necessary in order to create a dataset with a constraint of "better offer": if there are several lower quality offers before a better offer for the same patient, all those previous offers could be used to build a time target of better offer and including them all obviously contradicts any independence assumption.

On the experimental side, the improvements include other tests but also totally different approaches. As soon as more data is available, it would be interesting to compare the performances of the model with forward in time training set (donors arrived after the current offer). Carefully done, the comparison could somehow capture the best potential of the algorithm. To confront the performances of our algorithm, it would be valuable to try different methods: neural networks or machine learning approaches and stochastic simulations; on different datasets (e.g. U.S. data, synthetic data). There are several tuning possibilities from our work: we showed that cPRA was useful but not well calibrated for probability predictions. Assuming our model is perfect, we could fit a function of cPRA representing the actual needed probability on the validation set, to minimise the MSE or any relevant measure of performance, before validating on the test set. This would be an important contribution to the transplantation field too. A similar approach could be used to calibrate our current waiting-list variant of the algorithm and benefit from its good predictive power. If our algorithm was used in clinic, it would be interesting to evaluate the survival of patients who benefited from it, to detect any possible side effect of our methodology.

We want to give an idea of the potential of data from TQ for other research directions:

- study patients' decisions, assessing patients' preferences;
- study the current allocation policy and confront it to optimal policies, to support or criticise the current model.

For future research as well as for real-time clinical use of the algorithm, data should be taken care of. Data should be collected in such a way that it is easy to know:

- which patients were on the waiting-list at which time;
- which patients actually got a kidney offer at which time;
- which patients have a priority from which time to which time and when this priority is active or not for each offer.

These requirements are always hard to achieve in the medical world. For the past few years, the *Canadian Organ Replacement Register* reported increased under-reporting of data across Canada, and especially in Québec (CORR, 2015). Data collection and reporting are a challenge from the clinical level to the central level, also in order to encourage data scientists to get involved in medical projects.

BIBLIOGRAPHY

J. J. Abellán, C. Armero, D. Conesa, J. Pérez-Panadés, M. A. Martínez-Beneito, O. Zurriaga, M. J. García-Blasco, and H. Vanaclocha, "Predicting the behaviour of the renal transplant waiting list in the país valencià (spain) using simulation modeling," in *Simulation Conference, 2004. Proceedings of the 2004 Winter*, vol. 2. IEEE, 2004, pp. 1969–1974.

J.-H. Ahn and J. C. Hornberger, "Involving patients in the cadaveric kidney transplant allocation process: A decision-theoretic perspective," *Management Science*, vol. 42, no. 5, pp. 629–641, 1996.

A. Akl, A. M. Ismail, and M. Ghoneim, "Prediction of graft survival of living-donor kidney transplantation: nomograms or artificial neural networks?" *Transplantation*, vol. 86, no. 10, pp. 1401–1406, 2008.

O. Alagoz, L. M. Maillart, A. J. Schaefer, and M. S. Roberts, "Determining the acceptance of cadaveric livers using an implicit model of the waiting list," *Operations Research*, vol. 55, no. 1, pp. 24–36, 2007.

O. Alagoz, H. Hsu, A. J. Schaefer, and M. S. Roberts, "Markov decision processes: a tool for sequential decision making under uncertainty," *Medical Decision Making*, vol. 30, no. 4, pp. 474–483, 2010.

V. B. Ashby, A. B. Leichtman, M. A. Rees, P. X.-K. Song, M. Bray, W. Wang, and J. D. Kalbfleisch, "A kidney graft survival calculator that accounts for mismatches in age, sex, hla, and body size," *Clinical Journal of the American Society of Nephrology*, pp. CJN–09 330 916, 2017.

P. C. Austin and E. W. Steyerberg, "Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable," *BMC medical research methodology*, vol. 12, no. 1, p. 82, 2012.

D. A. Axelrod, M. A. Schnitzler, H. Xiao, W. Irish, E. Tuttle-Newhall, S.-H. Chang, B. L. Kasiske, T. Alhamad, and K. L. Lentine, "An economic assessment of contemporary kidney transplant practice," *American Journal of Transplantation*, 2018.

S. Bae, A. B. Massie, X. Luo, S. Anjum, N. M. Desai, and D. L. Segev, "Changes in discard rate after the introduction of the kidney donor profile index (kdpi)," *American Journal of Transplantation*, vol. 16, no. 7, pp. 2202–2207, 2016.

C. Bandi, N. Trichakis, and P. Vayanos, "Robust multiclass queuing theory for wait time estimation in resource allocation systems," *Management Science*, 2018.

M. Bendersky and I. David, "The full-information best-choice problem with uniform or gamma horizons," *Optimization*, vol. 65, no. 4, pp. 765–778, 2016.

——, "Deciding kidney-offer admissibility dependent on patients' lifetime failure rate," *European Journal of Operational Research*, vol. 251, no. 2, pp. 686–693, 2016.

W. M. Bennett and K. M. McEvoy, "A new system for kidney allocation: The devil is in the details," *Clinical Journal of the American Society of Nephrology*, vol. 6, no. 9, pp. 2308 – 2309, 2011.

D. Bertsimas, V. F. Farias, and N. Trichakis, "Fairness, efficiency, and flexibility in organ allocation for kidney transplantation," *Operations Research*, vol. 61, no. 1, pp. 73–87, 2013.

M. E. Brier, P. C. Ray, and J. B. Klein, "Prediction of delayed renal allograft function using an artificial neural network," *Nephrology Dialysis Transplantation*, vol. 18, no. 12, pp. 2655–2659, 2003.

Canadian Blood Services, "The facts about whole blood," https://blood.ca/en/blood/facts-about-whole-blood, 2018, accessed: 2018-05-28.

Y. H. Chun and R. T. Sumichrast, "A rank-based approach to the sequential selection and assignment problem," *European Journal of Operational Research*, vol. 174, no. 2, pp. 1338–1344, 2006.

CIHI, "Wait lists improving for some organs as deceased donor numbers jump in canada," https://www.cihi.ca/en/wait-lists-improving-for-some-organs-as-deceased-donor-numbers-jump-in-canada, 2017, accessed: 2018-05-14.

CORR, "Canadian organ replacement register, annual report: Treatment of end-stage organ failure in canada, 2004 to 2013," Canadian Institue for Health Information, Tech. Rep., 2015.

A. Dag, A. Oztekin, A. Yucel, S. Bulur, and F. M. Megahed, "Predicting heart transplantation outcomes through data analytics," *Decision Support Systems*, vol. 94, pp. 42–52, 2017.

G. M. Danovitch, *Handbook of kidney transplantation*. Lippincott Williams & Wilkins, 2009.

T. J. DiCiccio and B. Efron, "Bootstrap confidence intervals," *Statistical science*, pp. 189–212, 1996.

A. Edwards and G. Elwyn, *Shared decision-making in health care: Achieving evidence-based patient choice*. Oxford University Press, 2009.

Z. Erkin, M. D. Bailey, L. M. Maillart, A. J. Schaefer, and M. S. Roberts, "Eliciting patients' revealed preferences: an inverse markov decision process approach," *Decision Analysis*, vol. 7, no. 4, pp. 358–365, 2010.

D. Faraggi and R. Simon, "A neural network model for survival data," *Statistics in medicine*, vol. 14, no. 1, pp. 73–82, 1995.

B. E. Flores, "A pragmatic view of accuracy measurement in forecasting," *Omega*, vol. 14, no. 2, pp. 93–98, 1986.

J. S. Gill and M. Tonelli, "Penny wise, pound foolish? coverage limits on immunosuppression after kidney transplantation," *New England Journal of Medicine*, vol. 366, no. 7, pp. 586–589, 2012.

M. R. Gillick, "Re-engineering shared decision-making," *Journal of medical ethics*, pp. medethics–2014, 2015.

E. Gordon, Z. Butt, S. Jensen, A. Lok-Ming Lehr, J. Franklin, Y. Becker, L. Sherman, W. Chon, N. Beauvais, J. Hanneman *et al.*, "Opportunities for shared decision making in kidney transplantation," *American Journal of Transplantation*, vol. 13, no. 5, pp. 1149–1158, 2013.

S. Greenfield, S. Kaplan, and J. E. Ware, "Expanding patient involvement in care: effects on patient outcomes," *Annals of internal medicine*, vol. 102, no. 4, pp. 520–528, 1985.

S. Greenfield, S. H. Kaplan, J. E. Ware, E. M. Yano, and H. J. Frank, "Patients' participation in medical care," *Journal of general internal medicine*, vol. 3, no. 5, pp. 448–457, 1988.

C. Harvey, "The kidney transplant process model," in *Winter Simulation Conference (WSC), 2015*. IEEE, 2015, pp. 3176–3177.

C. Harvey and J. R. Thompson, "Exploring advantages in the waiting list for organ donations," in *Winter Simulation Conference (WSC), 2016*. IEEE, 2016, pp. 2006–2017.

E. Heaphy, D. Goldfarb, E. Poggio, L. Buccini, S. Flechner, and J. Schold, "The impact of deceased donor kidney risk significantly varies by recipient characteristics," *American Journal of Transplantation*, vol. 13, no. 4, pp. 1001–1011, 2013.

E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American statistical association*, vol. 53, no. 282, pp. 457–481, 1958.

A. Krasnosielska-Kobos, "Multiple-stopping problems with random horizon," *Optimization*, vol. 64, no. 7, pp. 1625–1645, 2015.

S. Krikov, A. Khan, B. C. Baird, L. L. Barenbaum, A. Leviatov, J. K. Koford, and A. S. Goldfarb-Rumyantzev, "Predicting kidney transplant survival using tree-based modeling," *ASAIO Journal*, vol. 53, no. 5, pp. 592–600, 2007.

A. Laupacis, P. Keown, N. Pus, H. Krueger, B. Ferguson, C. Wong, and N. Muirhead, "A study of the quality of life and cost-utility of renal transplantation," *Kidney international*, vol. 50, no. 1, pp. 235–242, 1996.

A. M. Law, *Simulation modeling and analysis*, 5th ed. McGraw-Hill International Edition, 2015.

J. Li, G. Serpen, S. Selman, M. Franchetti, M. Riesen, and C. Schneider, "Bayes net classifiers for prediction of renal graft status and survival period," *World Academy of Science, Engineering and Technology*, vol. 39, 2010.

M. Luck, T. Sylvain, H. Cardinal, A. Lodi, and Y. Bengio, "Deep learning for patient-specific kidney graft survival analysis," *arXiv preprint arXiv:1705.10245*, 2017.

B. J. Manns, D. C. Mendelssohn, and K. J. Taub, "The economics of end-stage renal disease care in canada: incentives and impact on delivery of care," *International journal of health care finance and economics*, vol. 7, no. 2-3, pp. 149–169, 2007.

A. B. Massie, J. Leanza, L. M. Fahmy, E. K. Chow, N. M. Desai, X. Luo, E. A. King, M. G. Bowring, and D. L. Segev, "A risk index for living donor kidney transplantation," *American Journal of Transplantation*, 2016.

H.-U. Meier-Kriesche, F. K. Port, A. O. Ojo, S. M. Rudich, J. A. Hanson, D. M. Cibrik, A. B. Leichtman, and B. Kaplan, "Effect of waiting time on renal transplant outcome," *Kidney international*, vol. 58, no. 3, pp. 1311–1317, 2000.

S. Méléard, *Modèles aléatoires en Ecologie et Evolution.* Springer, 2016.

G. E. Noether, "Note on the kolmogorov statistic in the discrete case," *Metrika*, vol. 7, no. 1, pp. 115–116, 1963.

M. J. Pencina and R. B. D'Agostino, "Overall c as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation," *Statistics in medicine*, vol. 23, no. 13, pp. 2109–2123, 2004.

M. Pérez-Ortiz, P. A. Gutiérrez, M. D. Ayllón-Terán, N. Heaton, R. Ciria, J. Briceño, and C. Hervás-Martínez, "Synthetic semi-supervised learning in imbalanced domains: Constructing a model for donor-recipient matching in liver transplantation," *Knowledge-Based Systems*, vol. 123, pp. 75–87, 2017.

A. N. Pettitt and M. A. Stephens, "The kolmogorov-smirnov goodness-of-fit statistic with discrete and grouped data," *Technometrics*, vol. 19, no. 2, pp. 205–210, 1977.

O. Procurement and T. Network, "A guide to calculating and interpreting the kidney donor profile index (kdpi)," 2016.

P. S. Rao, D. E. Schaubel, M. K. Guidinger, K. A. Andreoni, R. A. Wolfe, R. M. Merion, F. K. Port, and R. S. Sung, "A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index," *Transplantation*, vol. 88, no. 2, pp. 231–236, 2009.

F. Reinaldo, M. A. Rahman, C. F. Alves, A. Malucelli, and R. Camacho, "Machine learning support for kidney transplantation decision making," in *Proceedings of the International Symposium on Biocomputing*. ACM, 2010, p. 48.

R. M. Ripley, A. L. Harris, and L. Tarassenko, "Non-linear survival analysis using neural networks," *Statistics in medicine*, vol. 23, no. 5, pp. 825–842, 2004.

C. Rose, Y. Sun, E. Ferre, J. Gill, D. Landsberg, and J. Gill, "An examination of the application of the kidney donor risk index in british columbia," *Canadian journal of kidney health and disease*, vol. 5, p. 2054358118761052, 2018.

B. Sandıkçı, L. M. Maillart, A. J. Schaefer, and M. S. Roberts, "Alleviating the patient's price of privacy through a partially observable waiting list," *Management Science*, vol. 59, no. 8, pp. 1836–1854, 2013.

T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, and N. Khovanova, "Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation," *Biomedical Signal Processing and Control*, 2017.

R. L. Streit, *Poisson point processes: imaging, tracking, and sensing.* Springer Science & Business Media, 2010.

H. Tang, J. F. Hurdle, M. Poynton, C. Hunter, M. Tu, B. C. Baird, S. Krikov, and A. S. Goldfarb-Rumyantzev, "Validating prediction models of kidney transplant outcome using single center data," *ASAIO Journal*, vol. 57, no. 3, pp. 206–212, 2011.

Transplant Québec, "Procédure d'opération normalisée-attribution rénale," http://www.transplantquebec.ca/sites/default/files/att-pon-104_v5_ls.pdf, 2016, accessed: 2018-05-15.

——, "Statistiques officielles 2017," http://www.transplantquebec.ca/sites/default/files/statistiques_officielles_2017.pdf, 2018, accessed: 2018-05-15.

L. N. Wasserstein, "Markov processes over denumerable products of spaces describing large systems of automata," *Problems of Information Transmission*, vol. 5, no. 3, pp. 47–52, 1969.

A. Wey, N. Salkowski, W. K. Kremers, C. R. Schaffhausen, B. L. Kasiske, A. K. Israni, and J. J. Snyder, "A kidney offer acceptance decision tool to inform the decision to accept an offer or wait for a better kidney," *American Journal of Transplantation*, vol. 18, no. 4, pp. 897–906, 2018.

R. A. Wolfe, V. B. Ashby, E. L. Milford, A. O. Ojo, R. E. Ettenger, L. Y. Agodoa, P. J. Held, and F. K. Port, "Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant," *New England Journal of Medicine*, vol. 341, no. 23, pp. 1725–1730, 1999.

S. A. Zenios, G. M. Chertow, and L. M. Wein, "Dynamic allocation of kidneys to candidates on the transplant waiting list," *Operations Research*, vol. 48, no. 4, pp. 549–569, 2000.